March 6, 2013

TRANSACTION COST-BENEFIT ANALYSIS, WITH APPLICATIONS TO FINANCIAL REGULATION

D. Bruce Johnsen

Available at: https://works.bepress.com/d_bruce_johnsen/7/
TRANSACTION COST-BENEFIT ANALYSIS,
WITH APPLICATIONS TO FINANCIAL REGULATION

D. Bruce Johnsen
George Mason University School of Law
3301 N. Fairfax Drive
Arlington, VA 22201
703-978-7601
djohnsen@gmu.edu

© Preliminary draft. Do not cite or quote without the author’s permission.
TRANSACTION COST-BENEFIT ANALYSIS, WITH APPLICATIONS TO FINANCIAL REGULATION

Abstract

As Coase convincingly showed, transaction costs inhibit the ability of market participants to achieve first-best outcomes. This paper proposes a novel and relatively simple alternative to traditional cost-benefit analysis when regulated parties face sufficiently low transaction costs that they can bargain directly or rely on competitive markets to set efficient terms of trade. In these settings, the only informational burdens financial market regulators need bear to assess corrective rules is to identify the relevant parties, the “good” they hope to exchange, and the transaction costs that inhibit them from maximizing joint gains from trade. A rule is justified only if the regulator can show it is likely to reduce transaction costs. Facing lower transaction costs, the parties will adjust their arrangements to increase joint gains and social welfare. There is no need for the regulator to grapple with picking the appropriate discount rate or quantifying and weighing total costs and benefits. This is information the parties — the men and women “on the spot” — are best able to identify on their own.
TRANSACTION COST-BENEFIT ANALYSIS, WITH APPLICATIONS TO FINANCIAL REGULATION

D. Bruce Johnsen *

I. Introduction

Economists have struggled for decades with how to do reliable cost-benefit analysis. Yet during this time, federal law has increasingly required regulators to integrate cost-benefit analysis into the rule making process. The problem is that quantifying the total costs and benefits of any regulatory rule is exceedingly difficult, as is choosing the proper rate at which to discount future flows of costs and benefits. The Neoclassical economic model of market exchange illustrates the theoretical welfare effects of trade embedded in demand and supply, including marginal and total valuation, marginal and total cost, and gains from trade in the form of consumer and producer surplus. But the main positive scientific point of the model is not to monetize costs, benefits, or net gains from trade, or even to predict exact prices and quantities; it is to hypothesize an equilibrium and to predict the direction of market participants’ responses to outside shocks, so-called “comparative statics” or “marginal effects.” Being limited to ordinal measurement — that is, ranking rather than counting — predicting the direction of effects is about the best economists can hope for with any reliability.

In a given legal and regulatory setting, we can be confident market prices will approximate marginal benefits and costs and that all feasible gains from trade will be exhausted, but we lack the information necessary to quantify what is going on within or

---

* The author is Professor of Law at George Mason University School of Law. He holds his B.A., M.A., and Ph.D. (Economics) from the University of Washington and his J.D. from Emory Law School. He thanks Doug Ginsburg and George Priest for early encouragement on this project, Dan Monahan for diligent research assistance, and the Law and Economics Center and Mercatus Center at George Mason University for generous funding.

without the margin.\textsuperscript{2} This makes it extremely difficult as a policy matter to say anything reliable about whether one regulatory setting is likely to provide greater net benefits to society than another. This paper proposes a novel and relatively simple alternative to traditional cost-benefit analysis for those situations in which regulated parties face sufficiently low transaction costs that they can bargain directly or rely on competitive markets to set efficient terms of trade. I call it \textit{Transaction Cost-Benefit Analysis (TCBA)}.

In \textit{The Problem of Social Cost},\textsuperscript{3} Nobel Laureate Ronald Coase convincingly argued that if transaction costs were zero the institutional framework — the rule of liability, the form of property rights, the regulatory setting, the prevailing contract terms and business customs — would have no effect on resource allocation.\textsuperscript{4} In that case cost-benefit analysis would be irrelevant to promoting the public interest. In the real world, a world of costly transacting, the parties to an exchange will adjust their arrangements to maximize the joint gains from trade (social welfare) subject to the constraint transaction costs impose. Coase’s point was that it is methodologically meaningful to recognize transaction costs as a friction inhibiting their ability to do so.\textsuperscript{5} If we can identify how transaction costs differ across circumstances or over time, we should be able to predict or explain the parties’ choice of institutional framework. In light of the traditional economic justification for regulation, this means that every “market failure” provides the parties with a profit opportunity from choosing the institutional framework that minimizes

\textsuperscript{2} Even with terabytes of grocery store scanner data on prices and sales, it is impossible to estimate the demand for a particular product over more than a very limited range. Steve Tenn, \textit{Estimating Promotional Effects with Retailer-Level Scanner Data}, available at http://www.ftc.gov/be/workpapers/wp264.pdf

\textsuperscript{3} 3 Journal of Law and Economics 1 (1960).


\textsuperscript{5} Financial economists, financial market regulators, and financial market participants will be inclined to think of transaction costs as the costs of executing securities trades, consisting among other things of brokerage commissions, price impact, and the bid-ask spread. I use the term transaction costs far more broadly to include any of the costs that result from frictions in the process of market or nonmarket exchange.

The preoccupation of finance-types with the above conception of transaction costs can be traced back to an article widely cited in the finance literature by Harold Demsetz, \textit{The Cost of Transacting}, 82 Q.J.E. 33 (1968). Though Demsetz has his own quibbles with Coase’s analysis, he surely acknowledges transaction costs in the broadest sense as being a fruitful basis for explaining institutional arrangements.
transaction costs, all else being equal. As transaction costs go to zero, resource allocation approaches first-best.⁶

One important implication of Coase’s analysis is that government also faces transaction costs, and that the market cannot be said to “fail” in any meaningful sense — at least for the purpose of evaluating corrective regulation — unless government has comparative advantage at reducing transaction costs. In doubtful cases, many believe the burden of proof resides with the regulator. Recent federal case law agrees.

When transaction costs are modest, the only informational burden financial market regulators need bear to assess corrective rules is to identify the relevant parties, the “good” they hope to exchange, and the transaction costs that inhibit them from maximizing joint gains from trade. A rule is justified only if the regulator can show how it is likely to reduce transaction costs. The regulator need not grapple with the appropriate discount rate or quantify and weigh total costs and benefits. The parties — the men and women “on the spot”⁷ — are best able to identify and adjust to this information on their own.

Reagan-era reforms required executive agency personnel to perform enhanced regulatory cost-benefit analysis,⁸ and subsequent legislation required the U.S. Securities and Exchange Commission (SEC) to identify the likely effect of new rules on “competition, efficiency, and capital formation.”⁹ Since then financial market regulators’ use of traditional cost-benefit analysis has increased dramatically. Most of what has come out of the SEC has been understandably weak.¹⁰ If trained economists balk at accurately monetizing a rule’s aggregate costs and benefits, regulators with little or no economic training can scarcely be expected to do so. Transaction cost-benefit analysis dramatically reduces the regulator’s informational burden, leaving the transacting parties

---

⁹ Amendments to securities laws provided by NSMIA (1996) and Gramm-Leach-Bliley (1999). This requirement imposes the equivalent of a cost-benefit burden on the regulator.
¹⁰ See, for example, U.S. Chamber of Commerce v. SEC and Business Roundtable v SEC. See, also, SEC Rule 30(b)1-5 requiring investment managers to disclose their portfolio contents quarterly with a 60-day lag as opposed to every six months with the same lag. This disclosure requirement surely involves an opportunity cost to the manager from revealing what is arguably a trade secret.
to assess costs and benefits either explicitly or implicitly and to suffer the consequences accordingly. Transaction cost-benefit analysis requires the SEC and other financial market regulators to take seriously the field of transaction cost economics. A substantial body of careful theoretical and empirical scholarship exists to guide them. The U.S. Federal Trade Commission (FTC), and the U.S. Department of Justice, and federal courts have relied on this literature for decades.

Transaction cost-benefit analysis provides a viable alternative to traditional cost-benefit analysis where the parties bargain directly or transact in competitive markets. One clear example is regulation under the Investment Advisers Act (1940) and the Investment Company Act (1940). Advisers and their individual clients, whether individual or institutional, negotiate face-to-face. The same is true of broker-dealers and their clients under the Securities Exchange Act (1934). By all reasonable measures either these parties have the wherewithal and incentive to choose efficient contract terms or they are subject to intense competitive market forces that push them toward social

---

11 Nobel laureate Oliver Williamson probably coined the term “transaction cost economics” and was a leading member of a group of economists that revolutionized antitrust law by carefully applying it to various business practices, most successfully to vertical business arrangements. The Antitrust Division of the U.S. Department of Justice and the U.S. Federal Trade Commission routinely rely on scholarly work in this field. A search of the FTC website identified 242 references to Coase and 342 references to Williamson. A similar search of the U.S. Securities and Exchange Commission’s website identified only 4 references to Coase and 0 references to Williamson, despite the SEC’s authority over investment advisers and investment companies, which involve the same kind of vertical arrangements the FTC and DOJ have had to come to grips.


13 15 USC §80b
14 15 USC §80a
15 15 USC §78a
efficiency. Transaction costs may inhibit their ability to achieve a first-best outcome, and it is reasonable to suggest that SEC rule making can eliminate or at least ameliorate any remaining “inefficiencies.” This essay demonstrates the viability of TCBA to guide regulation under these acts, although it can be easily applied to a much wider range of activity in which transaction costs are moderate. My hope, is that the SEC will integrate TCBA into its tool-kit in keeping with a recent executive order instructing agencies to apply cost-benefit principles using the “best available techniques.”

Transaction cost-benefit analysis is a tractable and intellectually respectable framework within which to assess proposed corrective regulation in modest-transaction-cost settings. The informational burden it imposes on regulators is substantially less than traditional cost-benefit analysis. It limits the regulator’s inquiry to identifying the relevant parties, the good they hope to transact, the economic basis for their interaction, and the transaction costs that inhibit their exchange. In addition to being helpful to regulators sincerely focused on promoting investor welfare, TCBA also stands to provide a barrier to rash, misguided, or ill-motivated regulation and offers a viable constraint on regulatory capture or special interest rent seeking. It also mobilizes informed parties’ expertise and self-interest.

The goal for those charged with regulating in a modest-transaction-cost setting should be to eventually understand transaction costs sufficiently well that some system of property rights can emerge as a viable alternative to regulation. In that case the parties can be left to fend for themselves with the aid of standard private law rules. No doubt in many cases a plausible application of TCBA to financial markets will lead to the conclusion that there is nothing the regulator can do to reduce the parties’ cost of transacting and so any proposed regulation is impossible to justify.

Section II provides a brief history of federal regulators’ use of cost-benefit analysis and a quick look at the SEC’s use of cost-benefit analysis as revealed in three recent cases in the U.S. Court of Appeals for the District of Columbia Circuit. All of

17 What we know as private property rights is not necessarily a viable goal. Economists and other social scientists have shown that in many settings some form of common property, or even open access, provides parties with superior incentives to promote efficient resource allocation. The scholarly literature oftentimes shows that market participants in an open access or common property setting have evolved toward private property rights, but this outcome is far from inevitable.
these cases reject the SEC’s cost-benefit analysis as “arbitrary and capricious” under the Administrative Procedures Act (APA).\textsuperscript{18} There are recent signs of hope from the SEC, however. Its Division of Risk, Strategy, and Financial Innovation (RSFI) and the Office of the General Counsel (OGC) jointly produced a 2012 memo titled \textit{Current Guidance on Economic Analysis in SEC Rulemakings}, directing the staff to take cost-benefit analysis more seriously going forward. Section III provides an overview of traditional cost-benefit analysis, including a basic discussion of the underlying theory and methodology. It also takes a closer look at market failure as a justification for regulation. Section IV briefly examines the influence of transaction cost economics in antitrust law and policy, where it has generated a virtual revolution in regulation and jurisprudence. It then examines the SEC’s internal memo and provides a \textit{TCBA} analysis of the SEC’s traditional regulatory practice, with applications to various actions it has implemented under the Securities Exchange Act (SEA), the Investment Company Act (ICI), and the Investment Advisers Act (IAA). I specifically address market timing and mandatory standardized fee/expense reporting. Section V provides a summary and concluding remarks.

II. History, Current Practice, and SEC use of Cost-Benefit Analysis

This section provides background. Part A examines the history and current practice of cost-benefit analysis in federal regulation. Part B examines cost-benefit analysis at the SEC. It describes the likely sources of the SEC’s legal mandate to use it, as well as three recent cases in which the U.S. Circuit Court for the District of Columbia Circuit overturned SEC regulations for failure to conduct proper cost-benefit analysis. It then finishes with a discussion of the SEC’s response in what appears to be a sincere initiative to improve its practice, most notably the move to involve staff economists in the process earlier and more forcefully.

A. History and Current Practice of Cost-Benefit Analysis

It is difficult to pinpoint the origin of cost-benefit analysis, but it undoubtedly gained considerable traction with the rise of the administrative state starting in the U.S. with the New Deal.\textsuperscript{19} Cost-benefit analysis was apparently at issue in England during the 1930s debate over the Corn Laws.\textsuperscript{20} In the U.S., there is evidence the Army Corp of Engineers used it, at least informally, to evaluate various damn projects on the Snake and Columbia Rivers, also during the early 1930s.\textsuperscript{21} More formal use of cost-benefit analysis apparently began during the Johnson administration, with modestly increasing importance and sophistication during the Nixon, Ford, and Carter administrations.\textsuperscript{22} In 1980, President Carter signed the Paperwork Reduction Act into law.\textsuperscript{23} This statute created the Office of Information and Regulatory Affairs (OIRA) to “review and approve agency collections of information, including those related to regulations.”\textsuperscript{24} Shortly thereafter President Reagan put teeth into regulatory oversight with his Executive Order 12,291, mandating that all executive agencies perform cost-benefit analysis. Section 2 of the Order states that:

In promulgating new regulations, reviewing existing regulations, and developing legislative proposals concerning regulation, all agencies, to the extent permitted by law, shall adhere to the following requirements:

(a) Administrative decisions shall be based on adequate information concerning the need for and consequences of proposed government action;

\textsuperscript{20} Simon Lester 53 Va. J. Int'l L. Dig. 9, 11(2012).
\textsuperscript{22} See Jim Tozzi, \textit{OIRA’s Formative Years: The Historical Record of Centralized Regulatory Review Preceding OIRA’s Founding}, 63 Admin. L. Rev. 37 (2011).
(b) Regulatory action shall not be undertaken unless the potential benefits to society for the regulation outweigh the potential costs to society;

(c) Regulatory objectives shall be chosen to maximize the net benefits to society;

(d) Among alternative approaches to any given regulatory objective, the alternative involving the least net cost to society shall be chosen; and

(e) Agencies shall set regulatory priorities with the aim of maximizing the aggregate net benefits to society, taking into account the condition of the particular industries affected by regulations, the condition of the national economy, and other regulatory actions contemplated for the future.

The Order made OIRA responsible for analyzing regulations to ensure they plausibly maximize aggregate net benefits to society. It made provision for regulatory impact analysis and review of “major rules” and required each covered agency to publish biennially proposed regulatory agendas for the coming year. It made provision for the Director of the Office of Management and Budget to “[m]onitor agency compliance with the requirements of this Order and advise the President with respect to such compliance.” Based on its language, the Order arguably applied only to executive agencies.

Although OIRA’s early years were rocky, it eventually became a powerful though conspicuously obscure force on the federal regulatory landscape. One saving grace may have been President Clinton’s Executive Order 12,866 replacing Order 12,291. Order 12,866 maintains the substantive cost-benefit provisions of Order 12,291 but adds others. Notably, it frames the call for regulation in the language of market failure. Its Statement of Philosophy and Principles provides that

---

26 Gramm, at 29-30;
28 Executive Order No. 12,866, 58 FR 51735 (September, 1993).
Federal agencies should promulgate only such regulations as are required by law, are necessary to interpret the law, or are made necessary by compelling public need, such as material failures of private markets to protect or improve the health and safety of the public, the environment, or the well-being of the American people.

It imposes disclosure requirements on OIRA to ensure the transparency of its internal operations. It requires each agency to provide public notice of pending rules and to allow the public to submit comments for a period not less than 60 days. It makes the Vice President the principal adviser to the President on regulatory affairs and requires each agency, including independent agencies, to “prepare a Regulatory Plan (Plan) of the most important significant regulatory actions that the agency reasonably expects to issue in proposed or final form in that fiscal year or thereafter.” Order 12,866 remains in effect today, but in January 2011 President Obama reinforced it with Executive Order 13,563 requiring executive agencies to allow Internet submission of public comments, to provide for greater coordination with other agencies, to ensure scientific integrity, and to provide for retrospective analysis of existing rules.

B. Cost-Benefit Analysis at the SEC

As an independent agency, the SEC is arguably exempt from Executive Order 12,866. Some knowledgeable commentators believe the President has the power to bring independent agencies within its mandate, however, among them the American Bar Association, the U.S. Department of Justice, and Robert Hahn and Cass Sunstein. Hahn and Sunstein point out that the main difference between executive and independent agencies is that the President can remove executive agency appointees at will, while he

29 Gramm, at 29-30; Arbuckle, at 131-132.
30 Executive Order No. 12,866, 58 FR 51735 (September, 1993).
31 Executive Order No. 13,563, 76 FR 3821 (January 18, 2011).
can remove independent agency appointees only for cause — “inefficiency, neglect of duty, or malfeasance in office.” They make the following point:

What counts as “inefficiency”? What counts as “neglect of duty”? At a minimum, it seems sensible to say that the President is entitled to require the production of documents to ensure that agency officials are acting efficiently and in a way that does not show neglect. To this extent, procedural incorporation of the independent agencies — to require them to assess both costs and benefits — seems to be on firm ground. And it would be reasonable to go further. If agencies proceed when the benefits do not exceed the costs, they might reasonably be thought to be acting “inefficiently” (pun intended), and in a way that shows a neglect of duty, at least if the decision to proceed is not adequately explained.33

1. The Case Law

This is a plausible proposition, but with regard to the SEC’s use of cost-benefit analysis it may be moot. In 1996 Congress passed the National Securities Market Improvement Act (NSMIA),34 and in 1999 it passed the Gramm-Leach-Bliley Act (GLBA).35 These acts added roughly the following language to the SA (1933), the SEA (1934), the ICA (1940), and the IAA (1940):

(c) Consideration of promotion of efficiency, competition, and capital formation: . . . Whenever pursuant to this subchapter the Commission is engaged in rulemaking and is required to consider or determine whether an action is consistent with the public interest, the Commission shall also consider, in addition to the protection of investors, whether the action will promote efficiency, competition, and capital formation.

33 Id. at 1535.
34 Pub. L. 104–290, 110 Stat 3416
Does this language require the SEC to perform cost-benefit analysis of its proposed regulations? In three separate cases the U.S. Circuit Court for the D.C. Circuit recently found that it does, and in each case it found the SEC’s cost-benefit analysis deficient.

In *U.S. Chamber of Commerce v. SEC (Chamber I)*\(^{36}\) the Chamber sought review of the SEC’s *Investment Company Governance Rule*.\(^{37}\) Over the dissent of two commissioners, the *Rule* would have conditioned various exemptions that most mutual funds enjoy from provisions of the ICA on their having boards of directors with at least 75 percent outside directors and an independent chairman.\(^{39}\) It also would have required independent directors to meet at least once quarterly in executive session out of the presence of any interested director.\(^{40}\) The SEC considered the rule “necessary or appropriate in the public interest or for the protection of investors” owing to recent scandals involving late trading of mutual fund shares, inappropriate market timing of fund shares, and misuse of nonpublic information about fund portfolios. It viewed these actions as a serious breakdown in internal fund management controls and evidence that the funds were being used for the benefit of the management company or its employees, possibly including inside directors, rather than for the benefit of fund shareholders. In the SEC’s view, the proposed fund governance standards would put fund boards in a better position to demand that management adhere to the highest of compliance standards and to better oversee activities that involve inherent conflicts of interest between funds and their managers.

The Chamber challenged the *Rule* on two grounds. It argued that the SEC lacked substantive authority under the ICA to adopt the two conditions and that the SEC’s rulemaking process had violated the APA. Judge Ginsburg writing for the Court found

---

\(^{36}\) 412 F.3d 133 (2005).

\(^{37}\) The SEA, the ICA, and IAA all allow persons aggrieved by a final order of the Commission to obtain review of the order in the United States Court of Appeals for the circuit in which he resides or has his principal place of business or for the District of Columbia Circuit. SEA 15 U.S.C. § 78y(a); ICA Section 43(a), 15 U.S.C. § 78y(a), and IAA Section 13, 15 USC § 80b–13.

\(^{38}\) *Id.* at 56,759-60.

\(^{39}\) The ICA mandates that mutual funds have at least 40 percent outside directors. By ICA Rule 12(b)-1, the SEC conditioned various exemptions from prohibition on the use of fund assets to compensate for brokers’ sale of fund shares on the affected funds having a majority of outside directors. 17 CFR Parts 239, 240, 270 and 274. Release Nos. 33-7754; 34-42007; IC-24082; File No. S7-23-99.

\(^{40}\) The SEC never addressed the effect of this requirement on board collegiality.
that the SEC did have the substantive authority to condition access to various exemptions on compliance with its Investment Company Governance Rule. In his words,

…the Commission reasonably concluded that raising the minimum percentage of independent directors from 50% to 75% would “strengthen the hand of the independent directors when dealing with fund management, and may assure that independent directors maintain control of the board and its agenda.” Similarly, the Commission concluded that having an independent chairman would be beneficial because the chairman plays “an important role in setting the agenda of the board[,] ... in providing a check on the adviser, in negotiating the best deal for shareholders when considering the advisory contract, and in providing leadership to the board that focuses on the long-term interests of investors.” We have no basis upon which to second-guess that judgment.41

The Court disagreed with the SEC on two important points, however. First, it found that the SEC had failed to adequately consider the costs of the conditions it proposed and hence their likely effect on competition, efficiency, and capital formation. Although an empirical study is unnecessary, the SEC must nevertheless do its best to assess costs. The SEC claimed that the factual record on which it had based its ruling lacked a “reliable basis for determining how funds would choose to satisfy the [condition] and therefore it [was] difficult to determine the costs associated with electing independent directors.”42 The Court was unmoved, finding that uncertainty may limit what the Commission can do, but it does not excuse the Commission from its statutory obligation to do what it can to apprise itself — and hence the public and the Congress — of the economic consequences of a proposed regulation before it decides whether to adopt the

41 Chamber I at 141.
42 Chamber I at 143.
Accordingly, the Court found that the SEC had acted arbitrarily and capriciously in violation of the APA.

The Court also found that the SEC had failed to adequately consider the viable alternative rule of requiring funds to prominently state whether their board has an inside or independent chairman and to allow market forces to establish the best practice for each fund individually. Noting that the dissenting commissioners had strongly urged the SEC to seriously consider the disclosure alternative, the Court found that it had not adequately done so, that doing so would not have been frivolous or out of bounds, and that the SEC had therefore acted arbitrarily and capriciously in violation of the APA.

According to the Court, as a result of having neglected either to assess the cost of the rule or to show that it was the alternative involving the least net cost to society, the SEC inadequately considered the effects of its rule on competition, efficiency, and capital formation. Because these assessments are part of Executive Order 12,866, the Court’s opinion can be plausibly read to find that cost-benefit analysis, as that Order required, is implicitly subsumed within the SEC’s explicit statutory charge to consider a rule’s effect on competition, efficiency, and capital formation.

In American Equity v. SEC, the petitioner, American Equity Investment Life Insurance Company, sought the D.C. Circuit Court’s review of SEC Rule 151A under the SEA, promulgated in 2007. According to the rule, fixed index annuity contracts did not qualify for the Act’s exemption for “annuity contracts.” As an issuer of securities, American Equity therefore was subject to the Act’s registration and reporting requirements. With their inception in the mid-1990s, fixed index annuities had been regulated exclusively under state insurance laws, and by 2007 these contracts were

43 Chamber I at 144.
44 The aftermath of Chamber I is worth recounting. When the Court sent the matter back to the SEC, it had only eight days to reconsider the effect of the rule on costs before Chairman Donaldson’s term would expire. Having no time to add to the record, it relied on the existing record to reaffirm the rule. The Chamber again sought review. Since the SEC had already admitted that the existing record was inadequate to assess costs, the U.S. Circuit Court for the D.C. Circuit once again found it had violated the APA and sent the case back to the SEC for further consideration based on an enhanced record. Chamber of Commerce of U.S. v. SEC, 443 F. 3d 890 (2006) (Chamber II). The SEC then asked the economists in its Division of Risk, Strategy, and Financial Innovation to conduct its own empirical assessment and to survey the existing literature to determine the effect of outside directors and independent chairmen on fund performance. The results were unfavorable to the new chairman’s desire to pass the rule, so she instructed RSFI to file the studies. Shortly thereafter they were leaked to the press.[get the story and citations]
45 613 F.3d 166 (D.C. Cir. 2010).
selling at a yearly rate of roughly $24.8 billion with some $123 billion contracts outstanding.\textsuperscript{46} The SEC’s concern was that securities brokers subject to the SEA were not selling fixed index annuities owing to uncertainty over their legal status. By requiring disclosure, the rule would allow them to do so, thereby increasing competition and efficiency and benefitting the investing public. American Equity argued that the SEC had wrongfully excluded fixed index annuities from the Act’s annuity contract exemption and had failed to properly consider the Rule’s effect on competition, efficiency, and capital formation.

Writing for the Court, Judge Sentelle disagreed with the petitioner’s claim that the SEC had exceeded its authority in excluding fixed index annuities from the annuity contract exemption because they include elements of risk characteristic of a security rather than an insurance contract. But the Court agreed with the petitioner’s claim that the SEC ruling was arbitrary and capricious under the APA for failing to properly consider its effect on competition, efficiency, and capital formation. The thrust of the SEC’s rationale for the rule was that the absence of a clear rule identifying the regulatory status of fixed index annuities injected sufficient uncertainty into the market that competition, efficiency, and capital formation were being undermined. It was not enough for the SEC simply to declare that some rule is necessary. It must first establish a pre-rule benchmark for assessing competition, efficiency, and capital formation in the market for fixed index annuities and then identify the relative merits of the proposed rule in comparison to the benchmark. It had not done so. Accordingly, the Court found that the SEC lacked a reasoned basis for imposing the specific rule it had chosen.

Most recently, in Business Roundtable v. SEC\textsuperscript{47} the D.C. Circuit Court vacated SEA Rule 14a-11, known as the Proxy Access Rule.\textsuperscript{48} With modest limitations, the Rule would have required any company subject to the SEA, including investment companies, to add to their proxy materials the name of any person or persons nominated by a shareholder who has held at least 3% of the firm’s voting stock for at least three years. The effect of the rule would have been to allow qualified dissident shareholders a place on the ballot to elect the company’s board of directors. The SEC reasoned that the rule

\begin{footnotesize}
\textsuperscript{46} 613 F.3d 166, 170 (D.C. Cir. 2010).
\textsuperscript{47} 647 F.3d 1144 (2011).
\textsuperscript{48} 75 Fed. Reg. 56,668 (2010)
\end{footnotesize}
could create “potential benefits of improved board and company performance and shareholder value [that] justify [its] potential costs.”49 Over the dissent of two commissioners, it had rejected proposals to allow each company to choose whether to adopt and disclose the same requirements internally, arguing that “exclusive reliance on private ordering under State law would not be as effective and efficient” in helping shareholders exercise their control rights.50 It had also rejected a proposal to exclude investment companies from the rule.

One basis for the SEC’s ruling is that it would provide cost savings to dissident shareholders in the form of reduced printing, postage, and advertising expenses compared to a traditional proxy contest, in which dissident shareholders must bear these expenses out of their own pockets. Even though the SEC acknowledged the rule might have adverse effects on companies and their boards in terms of distraction from long-term strategic planning, it concluded that overall the rule would promote efficiency in the wider economy in excess of the associated costs. What is more, the SEC reasoned that any adverse effects on the board derive from longstanding state law proxy rules and not from enhanced proxy access.

Judge Ginsburg, again writing for the Court, flatly disagreed, finding the rule arbitrary and capricious in violation of the APA. In his words:

49 Id. at 56,761. The SEC’s failure may follow from an open letter published by Harvard Law Professor Lucien Bebchuck including supporting signatures from X law professors. The letter asserted that corporate managers were unduly accountable to shareholders and that some measure of added shareholder proxy access to correct the imbalance was necessary. The memo highlights the exact shortcomings of traditional CBA. It is surely true that complete director primacy, in which shareholders can do little to prevent managers from indulging their own preferences, is not socially optimal. Nor is complete shareholder empowerment. Directors must have some leeway to operate the firm without kibitzing by shareholders. Somewhere between these two extremes lies an optimal balance between director primacy and shareholder empowerment. But for Professor Bebchuck and his signatories to suggest they have some special insight to suggest that the balance should be readjusted in favor of greater shareholder empowerment is without any empirical or scientific foundation. If anything, one could easily have concluded that shareholders’ traditional recourse against management – the hostile takeover – had been unduly restricted by SEC rules. Conditional on this, perhaps greater proxy access is optimal, but why would it not be preferable simply to roll back excessively restrictive rule on the market for corporate control. See Henry G. Manne, “The Welfare of American Investors,” Wall Street Journal, available at http://online.wsj.com/article/SB115015714883578393.html. (or is it “The ‘Corporate Democracy’ Oxymoron”, Wall Street Journal, available at http://online.wsj.com/article/SB116768362469464053.html
50 Id. at 56,759-60.
[The SEC] inconsistently and opportunistically framed the costs and benefits of the rule; failed adequately to quantify the certain costs or to explain why those costs could not be quantified; neglected to support its predictive judgments; contradicted itself; and failed to respond to substantial problems raised by commenters. For these and other reasons, its decision to apply the rule to investment companies was also arbitrary.51

Among other things, the Court found that the SEC had arbitrarily dismissed an empirical study showing that when dissident directors win board seats in traditional proxy contests the firm underperforms its peers by 19 to 40 percent over the ensuing two years. It also faulted the SEC for attributing the costs of board distraction from enhanced proxy access as merely an incident of traditional state law proxy contest rules. Citing to Chamber I, the Court stated that “As we have said before, this type of reasoning, which fails to view a cost at the margin, is illogical and, in an economic analysis, unacceptable.”

The Court faulted the SEC for “ducking serious evaluation” of the costs arising from giving special interests such as union and public pensions greater access to corporate proxies, noting that these special interests might favor policies contrary to maximizing shareholder value such as increasing jobs.52 What is more, the Court found that the SEC had failed to properly estimate the number of proxy contests that would occur under the Rule because it had ignored the extent to which the Rule would merely displace the use of traditional proxy contests under 14a-8. It also erred by projecting frequent use of the Rule when assessing benefits but infrequent use when assessing costs.

Finally, the Court noted that the SEC had treated investment companies identically to operating companies, despite investment companies being subject to far more stringent shareholder empowerment rules under the ICA. How then could proxy access be assumed to provide the same benefits to investment companies as to operating companies? On the cost side, the Court pointed out that the SEC failed to consider the unique structure of investment company boards. Most investment companies are part of a family of funds with “unitary” boards, in which the same directors oversee the

51 See supra note 48, at 1148-49.
52 Id. at 1152.
operations of multiple funds. This structure provides efficiencies, for example by allowing issues involving different funds to be addressed in a single meeting. Allowing a dissident shareholder nominated under the Rule to have proxy access would eliminate this benefit and the expected benefits of the rule. The Court found the Rule arbitrary and capricious under the APA.

2. SEC Internal Initiatives

In March 2012 the SEC published an internal memo titled Current Guidance on Economic Analysis in SEC Rulemakings. The Division of Risk, Strategy, and Financial Innovation (RSFI) and the Office of the General Counsel (OGC) jointly authored the memo, which is truly remarkable for its conceptual clarity and apparent sincerity for integrating economic analysis into the rulemaking process. Although it correctly recognizes the difficulty of doing reliable cost-benefit analysis of financial regulations — especially when it comes to monetizing costs and benefits — it purports to take seriously the Chamber I court’s admonition that it has a “statutory obligation to determine as best it can the economic implications of [a proposed] rule.” Among other things, the memo suggests that RSFI economists be involved at the earliest stages of the rulemaking process, from the pre-proposal stage to the adopting stage. The RSFI economists should be primarily responsible for formulating the economic analysis. Although it finds that the SEC has no explicit statutory or executive duty to conduct cost-benefit analysis when it adopts a rule, it finds that doing so is simply a matter of good regulatory practice. Accordingly, it suggests that what had previously been treated in two separate sections of a proposing release — Cost-Benefit Analysis and Effect on Competition, Efficiency, and Capital Formation — should be integrated into a single Economic Analysis section, about which it has this to say:

54 Id. at 10, citing GAO Report No. 12-151 at 19 (citing GAO Report No. 08-32).
55 412 F.3d at 143.
It is widely recognized that the basic elements of a good regulatory economic analysis are: (1) a statement of the need for the proposed action; (2) the definition of a baseline against which to measure the likely economic consequences of the proposed regulation; (3) the identification of alternative regulatory approaches; and (4) an evaluation of the benefits and costs—both quantitative and qualitative—of the proposed action and the main alternatives identified by the analysis. As a general matter, every economic analysis in SEC rulemakings should include these elements, and the following guidance addresses ways to strengthen these aspects of our economic analyses.56

Among the benefits a rule might provide, the memo lists reduced incentive misalignment/reduced monitoring costs, lower cost of capital, better information sharing resulting in lower risk premiums and better allocation of capital, enhanced competition leading to reduced prices or higher quality, the avoidance of collective action problems, the avoidance of harmful transactions to reduce principal-agent problems such as excessive risk-taking or actions that are otherwise characterized as moral hazards, reduced transaction costs, and more efficient enforcement of SEC rules. Possible costs include the distributional and competitive effects of the rule, negative collateral consequences such as the potential misuse of newly created rights, and misallocation of resources resulting from regulatory arbitrage (race to the bottom).

The memo relies largely on the notion of market failure as the justification for regulation. It recognizes collective action problems as a primary reason market participants cannot resolve market failures through private ordering and lists the following as examples of market failure: “market power, externalities, principal-agent problems (such as economic conflicts of interest), and asymmetric information.”57 As I demonstrate below, there is good reason to be suspicious of market failure as the basis for regulation, including those market failures specifically listed. But nothing in the

56 See supra note 54, at 4.
57 Id. at 5.
statement is necessarily wrong, just possibly underspecified when full account is taken of the cost of transacting.

III. Overview of Traditional Cost-Benefit Analysis

This section lays the theoretical foundation for my analysis and application of TCBA in Section IV. Part A briefly describes the accepted Kaldor-Hicks rule for assessing welfare trade-offs in a cost-benefit framework. Part B examines the basic theory and methodology behind traditional cost-benefit analysis, which starts from the premise embodied in Executive Order 12,866 that regulation is justified when the market fails to efficiently allocate resources through private ordering. Part C re-examines market failure as a justification for corrective federal regulation. Early on, Frank Knight (1924) and later Ronald Coase (1958) pointed out some serious shortcomings in the concept of market failure as professed by A.C. Pigou. I show how focusing on transaction costs corrects these shortcomings for the purpose of cost-benefit analysis, at least for those situations in which the parties face only modest transaction costs.

A. Assessing Welfare Trade-offs

In 1896 Vilfredo Pareto proposed the criterion of Pareto optimality as a basis for making welfare trade-offs. A given allocation of resources is Pareto optimal if there is no reallocation that would improve one person’s welfare without reducing another’s. Though we will never know for certain, in a world of zero transaction costs voluntary market exchange would lead to Pareto optimality and market regulation would be unnecessary. Despite the appeal of relying exclusively on voluntary exchange to allocate resources, Pareto optimality is an unworkable standard for justifying regulation. Someone always loses from regulation, and the cost of finding the losers, divining their losses, and compensating them to assure that they would be no worse off is simply unworkable.

---

58 I beg knowledgeable readers’ indulgence for explaining what will be obvious in this section.
59 Much of the discussion in the sub-Part is based on Richard O. Zerbe, Jr., and Allen S. Bella, A Primer for Benefit-Cost Analysis (Edward Elgar, 2006).
The Kaldor-Hicks rule emerged in roughly 1939, and has become the default rule for assessing welfare trade-offs in the context of cost-benefit analysis. It says that a given re-allocation of resources is efficient if the winners could, in principle, fully compensate the losers and still improve their own welfare. It has given rise to the “potential compensation test.”

Any number of complications can arise in using the Kaldor-Hicks rule to assess welfare trade-offs. One criticism is that it requires the analyst to make inter-personal comparisons that have little or no foundation in economic theory. Do we accord weight in the welfare calculus to the value a murderer places on killing his victim? Another criticism is that it assumes transaction costs are zero. Yet another, and one that seems to have gotten a lot of attention, is that people’s valuations of a good appear to differ depending on whether they have it or not. Even ignoring income effects, there is considerable evidence that the minimum amount a given person is willing to accept (WTA) for a good he has is greater than the maximum amount he would be willing to pay (WTP) for the same good if he did not have it.

Richard Zerbe and Allen Bellas propose an alternative to the Kaldor-Hicks rule, which they label Kaldor-Hicks-Moral. Its main difference appears to be that it takes account of established legal rights, or even commonly held expectations, as a basis for establishing a status quo benchmark for assigning WTA and WTP valuations. It also excludes the valuations of people who violate these expectations based on a notion of legal standing.60

Kaldor-Hicks-Moral is an innovative and plausible refinement on the widely accepted Kaldor-Hicks rule that may address some of its many criticisms. Note that TCBA is subject to very few of these criticisms. It does not require the regulator to make interpersonal comparisons, it does not need to address whether a murderer’s valuation is...
legitimate, it surely does not assume away transaction costs, and it does not get bogged down in subtle and intractable distinctions between WTA and WTP. This is because it leaves these issues, to the extent they are even relevant, to the exchanging parties.

B. The Basic Theory Behind Traditional Cost-Benefit Analysis

Every undergraduate economics major learns that government regulation is justified when the market fails to allocate resources optimally. The standard story, following Adam Smith, is that people acting in their own self-interest will efficiently allocate resources only as long as they bear the full costs or capture the full benefits of their actions. When some costs or benefits are externalized to third parties the decision maker’s resource allocation decisions fall short of or exceed optimality and the market is said to fail. Some kind of government action in the public interest is then warranted to correct the market failure, the story goes, and can consist of quantity mandates or corrective taxation. Common examples of quantity mandates include speed limits and auto fuel efficiency standards, while the gasoline tax is an arguable example of corrective taxation.

1. The Neoclassical Model of Exchange

Market failure analysis refines the Neoclassical model of market exchange, shown in Figure 1. The per period rate of production and sales, Q, is measured on the horizontal axis, and the price in dollars per unit, P, is measured on the vertical axis. Line D shows consumer demand for the good, reflecting consumers’ aggregate marginal valuation (\(\sum MV_i\)), or willingness to pay, for each possible quantity. Line S shows producers’ supply of the good, roughly reflecting their aggregate marginal cost (\(\sum MV_i\)) for each possible quantity, with these costs equal to the value of productive inputs if deployed elsewhere. In a well-functioning competitive market with low transaction costs, the equilibrium price is \(P^*\) and the prevailing rate of production and sales is \(Q^*\). Consumers pay \(P^*\) for each

---

61 Mandatory disclosure, as under the SA, the SEA, and the IAA might be seen as a separate category of regulation. I treat it as a form of quantity mandate.
unit at \( Q^* \) units and make total expenditures equal to the associated rectangle in Figure 1. At the margin, for the \( Q^{\text{th}} \) unit of the good, consumer valuation is exactly equal to price, \( P^* \). Consumers are indifferent to whether they buy this unit or not, and so it generates no consumer surplus. Within the margin, however, consumers’ valuation of the good is hypothesized to exceed the price they pay. For \( Q^* \) units their total valuation, represented by the large trapezoid, exceeds their total expenditures by an amount equal to the upper shaded welfare triangle. This triangle is known as consumer surplus.

A similar story can be told for producers. For the \( Q^{\text{th}} \) unit they are indifferent to whether or not they supply it because \( P^* = MC \) for that unit. In total as a result of supplying \( Q^* \) units rather than none they earn total revenues of \( P^* \times Q^* \), exactly what consumers spend. Producers’ cost of supplying \( Q^* \) units is the area beneath \( MC \) from zero to \( Q^* \). They earn producer surplus equal to the lower shaded welfare triangle. Together, consumer and producer surplus reflect the hypothetical gains from trade, or social welfare. The language of regulatory cost-benefit analysis embedded in Executive Order 12,291 characterizes them as “aggregate net benefits to society.” The resulting allocation of resources is said to be efficient because no reallocation of resources would improve social welfare.

By far the most reliable use of the Neoclassical model is to predict the direction of the effect on price and quantity from outside shocks, such as the incidence of a tax, as illustrated in Figure 2. If the government imposes a tax of \( T \) dollars per unit on producers for all units traded they will naturally add \( T \) to the price they charge for any quantity supplied. This causes consumers to perceive the supply curve shifted up to \( S' = S + T \). Competitive forces will lead them to react to the higher price by reducing their consumption. As this happens, producers’ marginal production costs decline, as does the tax-inclusive price they require to supply a marginal unit of the good. Equilibrium is re-established at \( P^T \) and \( Q^T \), with \( P^T \times Q^T \) being consumers’ total expenditures. Consumer surplus is now the upper shaded welfare triangle in Figure 2. It has fallen by an amount equal to the upper elongated trapezoid. Producers’ total revenue is the same as \( P^T \times Q^T \), but they must remit \( T \times Q^T \) to the government. Their revenues net of the tax are equal to \( MC^T \times Q^T \), and their producer surplus is the lower welfare triangle in Figure 2. It has fallen by an amount equal to the lower elongated trapezoid.
Much of the loss in consumer and producer surplus is simply transferred to the government in the form of tax receipts, but there is a real efficiency loss equal to the dark-shaded triangle. This loss reflects gains from trade producers and consumers would enjoy absent the tax but that are lost owing to the its distorting effect on resource allocation. Economists refer to this as a deadweight loss.

The most reliable implications of the above model are that the price to consumers will rise, the net-of-tax price to producers will fall, and the quantity traded will fall, all as a result of the tax. The model has proven extremely successful in predicting the marginal effect of outside shocks. The predictions are testable against real world facts, and the underlying theory has gone largely unfalsified.

The reliability of the model’s implications falls off dramatically as we move beyond comparative statics. For example, economists’ ability to predict the magnitude of an increase in price to a new tax is imprecise and less precise the larger the tax. The same is true for predicting the magnitude of the reduction in the quantity traded. Among other reasons, this is because the economy is a noisy place with many things happening simultaneously. Having lots of data covering lots of variation to examine a fairly narrow range of activity can partially overcome this problem.

The model is even less reliable when it comes to generating welfare implications, that is, assessing how much better or worse off market participants are in total as a result of a tax or regulation. Economists hypothesize that total valuation is equal to the area under a demand curve we draw on a blackboard, but getting enough data to reliably estimate the demand curve is problematic. Not only is the economy a noisy place, but most of the variation we observe is in the narrow neighborhood of what economists characterize as equilibrium. In most settings we have no clue how much people would pay for the first hundred units of a good whose normal rate of consumption is in the millions. As a result we cannot be confident at what price the market demand curve intersects the vertical axis. Estimating consumer surplus — the difference between total

\[ \text{62} \text{ The resource distortion results from the excess of consumer marginal valuation over marginal cost between } Q^T \text{ and } Q^*. \]
\[ \text{63 In } \textit{Daubert v. Merrill Dow Pharmaceuticals}, 509 U.S. 579 (1993), \text{ the U.S. Supreme Court found that the falsifiability of the theory is one factor courts should consider in evaluating the admissibility of scientific expert testimony.} \]
valuation and total expenditures over the entire range of consumption possibilities — is identically problematic.

The same can be said on the producer side. Line S roughly reflects the combination of producers’ marginal costs, but the economic definition of cost is the value of the next best forgone opportunity. Opportunity costs are seldom observable in an objective way. Ordinarily they bear only a partial relationship to expenses and do not fully appear on balance sheets or income statements. Indeed, economists generally do not assert that market participants know the opportunity cost of their allocation decisions, only that they behave “as if” they know for the purpose of predicting their reaction to outside shocks.  

It is difficult to imagine how any real-world observation could refute statements about the magnitude of consumer or producer surplus derived from the Neoclassical model of exchange could be refuted by any real-world observation, which is the essence of scientific knowledge. This is not to say traditional cost-benefit analysis is hopeless. Over the years econometricians have made tremendous progress developing empirical methods to help see through the noise in the data and to disentangle the various factors that influence market outcomes, and much more complete data is now available. Even though sophisticated cost-benefit models may err in assessing social welfare differences under alternative regulatory regimes, in some cases there may be reason to believe the errors wash out and can therefore be confidently ignored. My point regarding TCBA is not that it precludes traditional cost-benefit analysis. Under some circumstances it might be a viable substitute, while in others it may be a helpful complement.

2. Market Failure

Assessing welfare when the market fails owing to either positive or negative externalities follows much the same economic logic as the tax incidence described above.

---

64 In a competitive market, those firms that happen to zig when they should zag will be eliminated from the system. Those remaining will appear to have chosen correctly even if they lacked the wherewithal to make an intelligent choice. Armen Alchian, Uncertainty, Evolution, and Economic Theory, 211 J.P.E. 58 (1950); Milton Friedman, The Methodology of Positive Economics (in Essays in Positive Economics, 1953).

Figure 3 illustrates the mechanics of market failure as traditionally conceived. Panel A demonstrates a negative externality, while Panel B demonstrates a positive externality. Line MPB reflects the marginal private benefits to a non-market participant from engaging in some activity. Because he captures all benefits from the activity there are no external benefits, and marginal private benefit is identical to marginal social benefit (MPB \equiv MSB). On the cost side the non-market participant’s private costs are given by MPC. As a rational maximizer he will engage in A^- units of the activity, where MPB = MPC. According to standard welfare analysis, at A^- he is said to capture total gains net of his own costs equal to the large shaded triangle.

The problem is that his decision about how much of the activity to engage in imposes an external cost on others. Because there is no market, the “victims” cannot communicate their preferences to him. He will do too much of the activity or, in the case of the positive externality shown in Panel B, too little. From society’s standpoint optimality occurs at A*. There, the actor’s marginal private cost, plus the external cost he imposes on others, defines marginal social cost (MPC + EC \equiv MSC), which should just equal the marginal social benefit. At A* social welfare would equal the large shaded trapezoid minus the smaller trapezoid, which includes the private costs to the decision maker plus the total external cost victims suffer from the market failure. As a result of being A^- rather than A*, social welfare falls short of the optimum by the dark shaded triangle, a deadweight loss.

Two possible solutions to the problem are regulation in the form of either a quantity mandate or a corrective tax. To implement a quantity mandate the regulator simply tells the actor to limit his activity to A*. The large trapezoid reflects his net welfare, now less than before the mandate by the small shaded triangle. Victims still bear external costs equal to the small elongated rectangle, but the resulting allocation of resources is nevertheless Kaldor-Hicks efficient. There is no reallocation of resources that will increase social welfare.

Like quantity mandates, corrective taxation imposes an information burden on the regulator but has the modest advantage of allowing the actor to decide the level of activity, subject only to having to pay the tax on each unit. The proper tax is said to equal EC. This shifts his marginal private cost from MPC to MPC + EC \equiv MSC. Acting
in his own self-interest he will naturally choose to engage in A* units of the activity, as in the case of quantity mandates. After remitting tax payments to the government equal to the small elongated rectangle, his net benefits are equal to the large triangle. In either case social welfare exceeds what it would be in the absence of regulation by an amount equal to the deadweight loss avoided. Note there is no social welfare difference between the quantity mandate and the corrective tax. The only difference is who suffers the burden of the externality. Under a quantity mandate the victims bear it. Under a corrective tax both the actor and the victims often bear it. With a tax the government collects revenue from the actor — equal to the small elongated rectangle — sufficient to compensate the victims, although compensation is often skipped because the transaction costs of following through on it are prohibitive.

C. A Closer Look at Market Failure

1. From Pigou to Knight to Coase

Early on Frank Knight questioned the accepted notion that market failure necessarily justifies regulation. In a response to Pigou’s identification of externalities as a rationale for corrective taxation, Knight re-examined Pigou’s example of two roads linking two cities. One road Pigou assumed to be slow but with sufficient capacity that it is never congested. The other road he assumed to be faster but subject to congestion. Under open access rights, and with excess demand, travelers join the fast road until it becomes so congested that the marginal traveler is indifferent between which road he chooses. This leads to overuse of the fast road because he neglects the congestion costs he imposes on his fellow travelers, a standard negative externality. What good does it do society to have a fast road if it does nothing to speed up travel?

Pigou proposed that an access tax to the fast road was necessary to correct overuse, but Knight showed that the optimal tax would be exactly the same as the profit

---

66 That is, the large welfare triangle minus the large elongated box is less than the smaller trapezoid minus the small elongated box.
67 Frank Knight, Some Fallacies in the Interpretation of Social Cost, 38 Q. J. Econ 582 (1924).
maximizing toll a private road owner would charge. From this, Knight concluded, it was not market failure that caused overuse of the fast road but Pigou’s unstated assumption that the road was unowned and therefore subject to open access. What is more, Knight made the important point that the social function of private property consists of the incentive it provides the road owner to gather the information necessary to identify the socially optimal toll.\footnote{Although Knight’s analysis focused on corrective taxation, it applies equally to quantity mandates.} It is entirely plausible that in many cases regulators lack the wherewithal or incentive to identify the optimal toll even if they know congestion when they see it.

Nearly 45 years later, Coase famously introduced transaction costs into the market failure debate, essentially helping to operationalize Knight’s insight about property rights. One plausible definition of transaction costs is that they consist of the costs of defining, enforcing, and exchanging property rights to myriad value flows.\footnote{But see Henry Smith, \textit{Property as the Law of Things}, 125 Harv. L. Rev. 1691 (2012).} Coase showed under the implausible condition of zero transaction costs that the form of property rights — and the institutional framework, more broadly — would have no effect on resource allocation. Where two or more people seek to use a scarce resource in mutually conflicting ways, the highest-valued use will inevitably prevail. If private property rights — including the right to exclude — reside with the low-value user, the high-value user will offer to pay the low-value user more to use the resource than the highest amount he would demand to relinquish it. The offer of payment becomes an opportunity cost to the low-value user from declining the offer. In his own self-interest he will relinquish the resource and accept the payment. The parties will jointly capture the gains from trade, sharing them in some \textit{Pareto} superior way. Alternatively, if property rights reside with the high-value user, the low-value user will be unable to offer a payment sufficient to compensate the high-value user to relinquish the resource. No trade occurs, but resource allocation is optimal and invariant to the initial assignment of rights.

2. The Parable of the Eggs
The opportunity to trade motivates specialization, which occurs when market participants invest to gain a comparative advantage in producing a narrower range of goods or services than what they consume. But in an important sense the whole point of trade is to internalize externalities as a matter of routine. Figure 4 illustrates the internalization of externalities. Imagine I own chickens that lay eggs, which I enjoy for breakfast. The horizontal axis shows the number of eggs I produce per week, while the vertical axis shows the value I place on eggs measured in dollars. The line MC shows my marginal cost of producing eggs. The more eggs I produce, the more it costs me in terms of forgone alternatives, such as using my limited resources to brew beer. The line MV\textsubscript{1} shows my marginal valuation. In the absence of trade I produce and eat E\textsuperscript{-} eggs per week, where my marginal valuation equals my marginal cost. The shaded triangle reflects my net gains.

Suppose my neighbor, Crusoe, looks over my fence one morning after catching the aroma of my Eggs Florentine and sees my chicken coop. He “really wants some eggs,” he tells me. My response is that I have no eggs to spare. Knowing I took the same undergraduate economics course he did, he refers me to Figure 4 and tells me I am injuring him by eating all E\textsuperscript{-} eggs per week. Is he right? Putting aside widely-shared moral assumptions about who has a right to what, the answer in one sense is yes. Line MV\textsubscript{1+2} reflects the aggregate value Crusoe and I place on eggs.\textsuperscript{71} The vertical distance between MV\textsubscript{1} and MV\textsubscript{1+2} is Crusoe’s marginal valuation. At zero eggs for Crusoe he values his first egg much more than I value my last egg at E\textsuperscript{-}. His unsatisfied demand, he insists, is a cost to him and, because he is a member of society, a cost to society as well. I have failed to properly allocate eggs. How dare I impose an externality on him?! For the benefit of society, he argues I should relinquish some of my eggs.

If my weekly egg production is fixed, and if this is a one-off problem, an omniscient regulator using cost-benefit analysis might conclude that forcing me to relinquish E\textsuperscript{-} minus E\textsuperscript{†} eggs would increase the excess of social benefits over social costs. With this reallocation we both place the same value, MV\textsuperscript{†}, on the marginal egg. Doing so is Kaldor-Hicks efficient. I lose net benefits equal to welfare triangle C, but

\textsuperscript{71} For any given quantity of eggs, the vertical difference between MV\textsubscript{1} and MV\textsubscript{1+2} reflects Crusoe’s valuation.
Crusoe picks up benefits of trapezoid D, for a net-net social gain equal to the difference, shaded trapezoid E.

A wise regulator is likely to recognize that there are relatively few one-off transactions between exchanging parties, that the number of one-off transactions is up to their discretion, and that in any event there are myriad private institutional arrangements (some yet to be invented) available to correct market failures. Two other things might come to the regulator’s attention. First, my egg production is not fixed and will decline if I must give up some of them for nothing. Second, and more important, the gains from voluntary trade exceed the gains from corrective regulation. Absent a forced transfer, and assuming my ownership of chickens and eggs is secure, Crusoe is willing to pay me to produce more eggs. If he offers P* I am happy to do so. Even though producing E* minus E¯ extra eggs imposes costs on me, the prospect of refusing P* for the extra eggs imposes an even bigger cost. At this price I am happy to produce the extra eggs and sell them to Crusoe. All things considered I am better off by welfare triangle A, my added producer surplus. Likewise, Crusoe is better off by welfare triangle B, his added consumer surplus. Our trade has internalized the externality from my underproduction of eggs and made us both better off.

That is not quite the end of the voluntary trade story. The next day Crusoe stops by to tell me I am keeping too many eggs for myself and therefore imposing a lingering externality on him. He shows me Figure 4 and explains that if I consume E¯ eggs while he consumes E* minus E¯ eggs, he values another egg based on the height of MV_{1+2} more than the net value I derive from my last egg, which is virtually zero. If I relinquish E¯ minus Eˆ eggs to him, he tells me, I can increase the gains from trade and should be willing to accept very little for doing so. Puzzled, I tell him I value the marginal egg at P*, the going market price. And, besides, I tell him he is willing to pay much more than P* to get it. He would rather get these eggs for free, but in the best interest of society he agrees to pay me P* per egg for E¯ minus Eˆ eggs. Presto! My producer surplus increases by triangle F and Crusoe’s consumer surplus increases by trapezoid G. We have fully corrected a market failure and captured all possible gains from trade in the

---

process. Social welfare is higher than under mandatory redistribution by trapezoid H minus triangle K.

Though this example proceeds stepwise to illustrate what ordinarily occurs spontaneously, it fairly makes the case that *trade solves the market failure arising from the absence of trade*. By definition, trade is the process by which market participants internalize externalities. They routinely encounter and resolve myriad market failures to the extent doing so makes them better off. Coase’s point was that the concept of transaction costs can be used both to operationalize the study of property rights (and other institutional arrangements) and to correct widespread misunderstanding about perceived market failure as the justification for corrective regulation.

To see how transaction costs fit into the model forget about eggs for the moment. Chickens stink! Figure 5 illustrates the Coasian point. Suppose the horizontal axis reflects the weekly labor effort in hours, L, I might deploy to abate the odor coming from my chicken coup. My costs and society’s are \( MPC = MSC \), with cost being the forgone value from eggs and beer I could have had instead. I dig a large pit in which to periodically deposit the chicken droppings and compost them with lyme and other stuff to reduce the odor. \( MPB \) gives the marginal private benefit to me from this labor effort. In my own self-interest I spend \( L^- \) hours per week at this project, where \( MPB \) intersects \( MPC \).

Now suppose Friday, my downwind neighbor, stops by. He is clearly unhappy. He holds his nose and points to my chicken coup. I guess he must dislike the odor, but because we do not speak the same language I am unsure what he wants me to do about it or why. After all, I am already taking what I consider to be appropriate measures to abate it. Unbeknownst to me he wants me to take further steps. In Figure 5 the labor effort I use to abate the odor benefits me, but it also provides Friday with an external benefit equal to \( EB \). The benefit to me, \( MPB \), plus \( EB \) added to it vertically, gives us the marginal social benefit, \( MSB \), from abating the odor. The external benefit appears to create a market failure because it means I do too little abatement. If I could be made magically to bear this additional cost I would exert additional labor effort up to \( L^* \), where \( MSB = MPC \). This would be in society’s best interest, but since I cannot communicate with Friday we are unable to work out a deal.
Suppose I can hire an interpreter at a cost reflecting the value of his forgone opportunities represented as IC. If we add IC to MPC we arrive at MSCₚ, the marginal social cost of using my labor and the interpreter’s labor to privately abate the odor in a way that accounts for the benefits to both Friday and to me. As shown, for any amount of labor IC exceeds EB, so that equating MSCₚ and MSB leads me to spend only L⁻ hours of labor per week to abate the odor. The cost of hiring an interpreter, IC, is a transaction cost, and is a real social cost to be avoided unless the associated benefits are even greater. Under the circumstances they are not. Recognizing that the absence of a benefit is a cost, for illustrative purposes we can redraw EB as the external cost, EC, resulting from changing my abating labor in the wrong direction. It is better for society by IC minus EC times L⁻ plus the small shaded triangle if I refrain from using the interpreter. Why would society spend one dollar to save 90 cents? Just because it would be nice if I were to magically consider EB in my decision calculus does not make my failure to do so a problem worth correcting. Because social welfare will fall from incurring the transaction cost necessary to avoid the market failure, it is no market failure in any meaningful sense.

Doing nothing is optimal unless the regulator has comparative advantage in reducing transaction costs. The regulator might mandate that I expand my abatement efforts to L*, but this raises the informational problem Knight identified. How does the regulator know L*? In addition, ongoing enforcement will be costly because I am disinclined to work for the benefit of others unless compensated for doing so. Only if the regulator’s cost per unit of my labor to hire and pay an interpreter is less than EB is it in society’s interest to regulate. Only then can we characterize my failure to account for EB as a true market failure requiring government regulation.

Suppose the government has comparative advantage in reducing transaction costs through regulation, perhaps by mandating education to increase English literacy. This reduces my transaction costs of negotiating with Friday to GC. If we add GC to MSC we arrive at MSCₚ. If GC is less than EB I naturally increase my labor effort from L⁻ to L† because Friday will pay me something to do so. We will both capture gains from trade.

73 Assume for convenience that IC increases with my labor inputs, rather than being a one-time cost.
Note that in the case of positive spillovers, all else being equal, the expansion of mutually beneficial trade is a good proxy for reduced transaction costs and increased social welfare. Note also that I am still unable to capture all the external benefits. There is a lingering spillover equal to the value of $\text{MSC}_G$ minus $\text{MSC}_P$ between $L^\dagger$ and $L^*$, but only in the sense that we can imagine a more perfect world rather than in the sense that society’s resources would be well spent getting us there.\(^{74}\) Similarly, in the case of negative spillovers a reduction in the offending activity is a good proxy for reduced transaction costs.\(^{75}\)

The moral of the parable is that market participants come together to capture the gains from trade. They are routinely willing and able to internalize externalities to avert market failure, even though doing so requires them to incur transaction costs. And they do so to satisfy their own self-interest rather than out of a desire to improve social welfare or adhere to regulatory edicts. Their joint, ongoing interest is to continue innovating to further reduce transaction cost lest money be left on the table. This is the driving force behind much financial engineering subject to SEC regulation. None of this is to suggest that corrective regulation is unwarranted, only that sensible regulation must take account of the transaction costs that hinder the parties from increasing the gains from trade through private ordering. According to TCBA, if the SEC is unable to clearly make this case it should refrain from further regulation lest it do more harm than good.

IV. Applications of Transaction Cost Benefit Analysis

Part A briefly reviews the influence of transaction cost economics in the field of antitrust law and policy. Part B examines the SEC’s internal memo and offers a few critical thoughts to fill gaps and correct various omissions and misconceptions. Part C evaluates the SEC’s traditional use of economic theory and applies TCBA to past and future SEC rule making. The analysis focuses on the regulation of investment advisers to

---

\(^{74}\) See Harold Demsetz, Information and Efficiency: Another Viewpoint, 12 J. Law & Econ. 1 (1969). In his words: “The view that now pervades much public policy economics implicitly presents the relevant choice as between an ideal norm and an existing ‘imperfect’ institutional arrangement. This nirvana approach differs considerably from a comparative institution approach in which the relevant choice is between alternative real institutional arrangements.”

\(^{75}\) These observations mirror Robert Bork’s assertion that business practices that increase output are likely to increase consumer welfare. See Robert Bork, The Antitrust Paradox (1978).
mutual funds under the ICA. The relations between advisers and their client, whether retail or institutional, are vertical relationships of the principal-agent form and suffer from the same inherent conflicts of interest as vertically situated parties in the antitrust setting. Transaction cost economics applies equally well to conflicts of interest between advisers and their clients.

A. The Role of Transaction Cost Economics in Antitrust Law and Policy

For many years the FTC’s Bureau of Competition, the DOJ’s Antitrust Division, and federal courts treated all manner of vertical arrangements between manufacturers and their distributors as presumptively unlawful. With the application of standard industrial organization economics, this began to change but gained increasing momentum with the advent of transaction cost economics. As the law now stands, for the most part, vertical arrangements are no longer presumptively unlawful. This part provides a brief history of the role industrial organization economics has played in the evolution of antitrust law and the subsequent role transaction cost economics played in revolutionizing how scholars, judges, and regulators think about competition policy.

Perhaps owing to a growing public fear of trusts and other large business combinations, in 1890 Congress passed one of the first major pieces of federal legislation, the Sherman Antitrust Act.\(^{76}\) The Act outlawed “[e]very contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States . . .” and provided that “[e]very person who shall monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize any part of the trade or commerce among the several States, or with foreign nations, shall be deemed guilty of a felony.” The Act provided a private right of action, and eventually the Clayton Act (1914) added treble compensatory damages for successful claimants.\(^{77}\) By making restraints of trade illegal and punishable as felonies, the Act dramatically departed from the common law, which had always treated naked trade restraints as simply \textit{nudum pactum} — having no force of law.

\(^{76}\) Ch. 647, 26 Stat. 209, 15 U.S.C. §§ 1–7

Cases under the Act soon found their way into the federal courts, including the U.S. Supreme Court. In 1911, Justice White wrote a seminal opinion in *U.S. v. Standard Oil of New Jersey* laying out the legal framework for judging claims under the Sherman Act.\(^\text{78}\) Trade restraints broadly construed were to be treated under the *rule of reason*, with courts’ factual inquiries focusing on the purpose or *intent* of the restraint, the participants’ *power* to restrain trade or commerce, and the likely *effect* of the restraint on trade or commerce. Most notably, White identified a category of restraints so unlikely to survive this factual inquiry that they would be held unreasonable as a matter of law — or, eventually, as unreasonable or illegal *per se*. By way of example, defendants who admitted to fixing prices but who attempted to show that their prices had been reasonable and therefore lawful under the Act were quickly met with summary judgment in favor of the government or plaintiff.\(^\text{79}\)

For decades following the Standard Oil decision courts grappled with how to apply the statute’s language to all manner of business practices, both vertical and horizontal.\(^\text{80}\) It is fair to say the case law was fitful and inconsistent, among other reasons because different courts used different standards for assessing the negative effects of the restraint at issue. Some courts focused on the effect of the restraint on “competition,” with this term being variously interpreted either as the competitive process or as the structure of the industry in terms of the number and market shares of rival firms. Other courts focused on goals such as “the preservation of small dealers and worthy men,” no doubt reflecting the strong populist sentiments of many who had favored passage of the Act in 1890. At times courts focused on exclusionary practices. At still other times they focused, and focused increasingly, on consumer welfare. Within this chaotic setting many business practices in any way out of the ordinary were condemned as restraints of

---

\(^\text{78}\) 221 U.S. 1 (1911)

\(^\text{79}\) See *U.S. v. Trenton Potteries*, 273 U.S. 392 (1927) (an agreement to fix and maintain uniform prices violates the Sherman Act whether the prices in themselves were reasonable or unreasonable). This case laid the foundation for the Court’s adoption of the category of restraints that are unreasonable *per se* in *U.S. v. Socony-Vacuum*, 310 U.S. 150 (1940).

\(^\text{80}\) A vertical arrangement is one between different firms in the chain of production and distribution. Examples include agreements between a manufacturer and its distributors or retailers requiring exclusive dealing, exclusive territories, requirements contracts, deferred rebates, product bundling (tie-ins), resale price maintenance, customer restrictions, vertical mergers, etc. Horizontal arrangements are those between rival firms at a given level of production or distribution. Examples include horizontal price fixing, horizontal division of territories, and horizontal mergers.
trade or attempts to monopolize. Many businessmen who engaged in them were fined or imprisoned based on muddled economic theories of how the law should be applied that have since been discredited. Ronald Coase best captured the state of economic understanding during this period in the following:

If an economist finds something — a business practice of one sort or another — that he does not understand, he looks for a monopoly explanation. And as in this field we are very ignorant, the number of [misunderstood] practices tends to be very large, and the reliance on a monopoly explanation, frequent.81

Among the practices economists, antitrust regulators, and federal courts widely condemned were the large number of vertical contractual arrangements negotiated between manufacturers and their distributors or retailers. They included exclusive dealing, in which a retailer agrees to sell the manufacturer’s product and to refrain from selling rival manufacturers’ products in the same product category.82 They included exclusive territories, in which a distributor agrees to sell only to customers within a specified geographic territory.83 They included both maximum and minimum resale price maintenance, wherein distributors or retailers agree to adhere to the maximum or minimum prices the manufacturer prescribed. They also included product bundling in the form of so-called “tie-in” sales, according to which a distributor or retailer agrees to buy the second of the manufacturer’s products if it buys the first. Bundling might also come in the form of a distributor’s or retailer’s agreement to buy and resell a manufacturer’s entire line of products or none at all. Finally, they included vertical mergers between manufactures and their distributors or retailers.

The early economic theory leading courts to condemn vertical arrangements, often as unreasonable per se, roughly reflects Coase’s observation above. The theory

was that a manufacturer who possessed market power — the ability to raise price above marginal cost without suffering a complete loss of sales — could extend and enhance that market power through its contractual relations with distributors or retailers, possibly by excluding its rivals from access to retail distribution. As the economics of industrial organization advanced, the theory came under ever-greater scrutiny and was eventually rejected as implausible in all but the rarest of cases.84 Any manufacturer that enjoys market power can capture all the associated benefits in the price it charges retailers for its product, and it cannot benefit by further vertical extension.

Rather than vertical arrangements extending market power, economists began to realize that the absence of downstream vertical restrictions would lead to inefficiencies in the form of “double marginalization,” depicted in Figure 6. Line D shows the demand a manufacturer possessing market power perceives for its product. Starting at any arbitrary price and quantity, the manufacturer can raise price without losing all sales. It is therefore a price searcher in the sense that it must discover the profit-maximizing price. Under the constraint that it has to set the same price for all units it sells, to sell an additional unit it must reduce price on all infra-marginal units as well. As a result, the manufacturer’s marginal revenue is equal to MR, which lies below consumers’ marginal valuation imbedded in D. Its choice of output and price creates a positive externality for which it may not account in deciding how much to produce and what price to charge. Acting in its own self interest the manufacturer will set output at $Q^*$ and price at $P^*$ to maximize monopoly profits (the large shaded rectangle), whereas $Q^*$ and $P^*$ would maximize social welfare. The shaded triangle identifies the social welfare loss from market power, which is a deadweight loss and the primary concern of antitrust regulators and courts.

Although Justice White’s opinion in Standard Oil identified the power to restrain trade as a relevant factor under a rule of reason analysis, antitrust regulators and courts now tolerate market power, unless extreme, for various good reasons.85 The question is

---

85 One good reason to tolerate market power is that it is very often the result of productive efficiency resulting from economies of scale. Breaking efficient large firms into inefficient small firms with higher production costs would more than outweigh any welfare loss from market power.
how much market power is too much, and what regulators and courts can effectively do about it, all things considered.

A manufacturer with market power faces a problem if it sells its product to retailers unconstrained at $P^-$. By definition, the retailer is likely to have market power in the same product. The demand it faces under plausible assumptions is $dd$, its marginal revenue is $rr$, and its marginal cost is the wholesale price it pays, $P^-$, plus the marginal cost of retailing, $MC_R$. Acting unconstrained, the retailer will sell $Q^*$ units of the good at a price of $P^*$, where its marginal revenue equals its marginal cost. The manufacturer’s monopoly profits fall (equal to the long thin shaded rectangle). The welfare losses from market power are now much larger, equal to the large shaded triangle.

The problem of double marginalization arises because the manufacturer and retailer face a principal-agent problem in which an inherent conflict of interest is ever present. The manufacturer has a strong interest to avoid the conflict by constraining its retailers to prevent double marginalization, and it clearly has no interest in sharing its market power with retailers.

If vertical arrangements do not allow manufacturers profitably to extend their market power to the retail level, why do manufacturers use them? Enter transaction cost economics. Following Coase’s publication of the *Problem of Social Costs*, economists began developing explanations for how various vertical arrangements reduce transaction costs. The effect was soon felt in antitrust case law. In *Continental TV v. GTE Sylvania* (1977) the Court found that manufacturer-imposed exclusive territories would no longer be considered unreasonable per se. The facts showed that GTE had been one of many television manufacturers competing in Northern California, apparently vigorously. After contracting with Continental TV and other retailers to establish exclusive territories within which its retailers could sell its televisions alongside other manufacturers’ televisions, GTE’s market share soon rose from roughly one percent to close to five percent. When it found Continental selling its televisions outside its San Francisco territory, GTE terminated the contract. The Court, recognizing the inherent conflict of interest between GTE and its retailers, found that the territorial restrictions induced GTE

---

86 433 U.S. at 41.
87 *Id.* at 38. If the territorial restrictions were designed to extend GTE’s market power to the retail level its market share would have declined over time rather than risen.
retailers to aggressively sell its televisions against non-GTE retailers (interbrand competition) but not against other GTE retailers (intra-brand competition). The arrangement reduced transaction costs by freeing GTE from having to carefully monitor its retailers’ sales and promotional efforts, which among other things probably included costly efforts to explain to consumers the relative merits of GTE televisions compared to rival brands.

Even before publication of Robert Bork’s influential treatise *The Antitrust Paradox: A Policy at War with Itself*, economists and antitrust regulators had begun to develop efficiency explanations for vertical arrangements based on transaction costs. Bork easily showed that virtually all contracts restrain trade. If I agree to sell a dozen eggs to Crusoe I cannot sell the same eggs to Friday. The Sherman Act’s prohibition of “all contracts, combinations, and conspiracies in restraint of trade” could not, therefore, be taken literally. Bork convincingly argued that a unitary goal for antitrust is necessary to parse good restraints from those that should be condemned. Promoting “consumer welfare” was, in Bork’s view, the exclusive goal of antitrust law. Among other things, he proposed a tractable rule of thumb for regulators and courts to follow in reviewing or deciding cases. In general, practices likely to increase the output of the good being transacted benefit consumers, while practices likely to reduce output harm consumers. In large part, the regulator or court need only divine the effect of a business practice on output to decide whether or not to condemn it.

Bork’s treatise, together with the ever-expanding body of scholarship in transaction cost economics, is responsible for one of the most dramatic reversals in judicial decision making imaginable. It was not long before courts began taking other vertical arrangements out of the *per se* category. This includes exclusive dealing, which likely reduces transaction costs by preventing retailers from free riding on the manufacturer’s promotional efforts getting the customer into the store and then recommending a lower-priced knock-off product instead. A variant of this bait-and-switch tactic has occurred in the market for branded gasoline, when branded retailers sell cheap off-brand gasoline out of pumps carrying their major supplier’s brand name. See D.

---

89 The effect of a business practice on output is a crude proxy for gains from trade, or social welfare. It might be in some cases that output falls but that the quality of the good rises by enough to more than offset the output effect, leaving consumer welfare higher than otherwise.
90 A variant of this bait-and-switch tactic has occurred in the market for branded gasoline, when branded retailers sell cheap off-brand gasoline out of pumps carrying their major supplier’s brand name. See D.
contracts. Eventually, it also came to include maximum resale price maintenance (which
directly avoids double marginalization) and even minimum resale price maintenance in
the fairly recent U.S. Supreme Court decision Leegin v. KSPS (2007).\textsuperscript{91} Especially where
the manufacturer has small market share, minimum resale price maintenance allows the
manufacturer to avoid monitoring its retailers to ensure they provide valuable point-of-
sale services and other non-contractible inputs to prospective buyers, all of which reduces
transaction costs. Rather than lowering price to compete, minimum resale price
maintenance encourages retailers to compete by providing these non-contractible inputs,
with less need for costly manufacturer oversight. Transaction costs fall as a result.

The FTC and DOJ have virtually abandoned review of vertical mergers, and now
focus most of their attention on horizontal restraints of trade and horizontal mergers.
Transaction cost economics has even profoundly influenced the law regarding horizontal
restraints of trade has.\textsuperscript{92}

B. Transaction Cost Economic Analysis of the SEC’s Internal Memo

The SEC’s \textit{Current Guidance} memo demonstrates the insight and scholarly
integrity of the economists who helped compose it. The Division of Risk, Strategy, and
Financial Innovation — and its predecessor, the Office of Economic Analysis — now
hosts, and has routinely hosted over the years, some of the best academic economists in
the field. Many have gone there to visit from top universities to take advantage of the
inside perspective and access to data it provides. Yet, until recently, the SEC has shown
little ongoing inclination to integrate this valuable resource into the institutional fabric of
its rulemaking process. Doing so does not require it to re-invent the wheel. It can easily
look to templates developed at the FTC and DOJ and adapt them to its particular
circumstances.

\begin{thebibliography}{99}
  \bibitem{92} 551 U.S. 877 (2007).
  \bibitem{92} See Broadcast Music, Inc. v. Columbia Broadcasting System, Inc., 441 U.S. 1 (1979);
\end{thebibliography}
Although the SEC’s *Current Guidance* memo briefly addresses the need to identify the “goals” of regulatory action,\(^93\) it must address this issue foremost if it is to rely on any form of cost-benefit analysis to assist in rulemaking. From their inception, all four of the major securities laws the SEC administers contained the very broad mandate to regulate “as necessary and appropriate in the public interest or for the protection of investors.”\(^94\) The NSMIA and Gramm-Leach-Bliley Act added the promotion of “competition, efficiency, and capital formation” to the SEC’s charge. This is more than a handful. As Robert Bork made clear in *The Antitrust Paradox*, it makes no sense to talk about multiple maximands — competition as process, competition as structure, preservation of small dealers and worth men, etc. — unless they will never come in conflict in a way that requires trade-offs.\(^95\) Trade-offs are inevitable. Consistent with the language of the statutes, and following Bork’s template, the SEC must identify and follow an overall maximand to balance the offsetting effects of a regulation on the public interest, investor protection, and the promotion of competition, efficiency, and capital formation\(^96\)

---

\(^93\) See supra note ??, at ??.


\(^95\) Investor protection is generally in the public interest, but not always. An exception might occur when we protect investors so much that they lack the good sense to come in from the rain. If so, moral hazard might lead them to neglect taking modest precautions whose absence imposes costs on others, such as by triggering financial contagion.

\(^96\) One relevant point to consider along the way is that the “public” whose interests are to be served consists of all financial market investors strictly *qua* financial market investors. There is little the SEC can do to further the interests of motorists or duck hunters, or consumers of minerals that originate in the Congo, as such, even though they may also be financial market investors. This point may strike readers as gratuitous. I refer them to various SEC rules that appear well outside of its mission. One recent example is the SEC’s final rule on conflict minerals. 17 CFR PARTS 240 and 249b, Release No. 34-67716; File No. S7-40-10. In its defense, Congress required the SEC to fashion some rule requiring certain companies to disclose their use of conflict minerals — originating in the Republic of the Congo — that include tantalum, tin, gold, or tungsten if those minerals are “necessary to the functionality or production of a product” manufactured by those companies. See Section 1502 of the Dodd-Frank Wall Street Reform and Consumer Protection Act (2010) H.R. 4173-840. Yet the U.S. Chamber of Commerce has filed a challenge to the rule the SEC actually fashioned. See *Andrew Harris, SEC Sued Over ‘Conflict Minerals’ Rule by Business Groups*, Bloomberg, Oct 22, 2012. This simply underscores the importance of identifying a clear objective for SEC financial regulation and a reliable body of economic theory to advance that objective in the rulemaking process. Doing so will allow the SEC to resist Congressional pressures to act outside of its mission. Get other examples of the SEC, on its own volition, regulating the substance of market exchange. The proxy access rule, discussed infra.
What about the welfare of the investment bankers, broker dealers, investment advisers, and many other intermediaries that serve investors? Imbedded in much of the SEC’s economic rationale for regulation has long been the “fixed-pie” assumption that when one party gains from market exchange — market insiders such as investment bankers, broker dealers, and investment advisers are the usual scapegoats — other parties necessarily lose, especially small investors.97 Aside from being a widely discredited vestige of Marxist ideology, the fixed-pie rationale is incapable of identifying the equilibrium conditions necessary to serve as an analytical foundation for any kind of cost-benefit analysis. It is devoid of operational content, at once leaving the SEC unconstrained in what and how it can regulate and unprotected from outside influences.

I have already made the point that market exchange tends toward an expanding pie, and that if not the cost of transacting is a methodologically meaningful culprit. If transaction costs were zero, the interests of those that serve the investing public would in turn be served because investors would pay them to do so. Because transaction costs are not zero the SEC may be required, as necessary or appropriate in the public interest, to consider market intermediaries’ interests and to trade them off against the interests of the investing public. This strikes me as a good explanation for Congress’s mandate that the SEC should also consider whether its regulations “will promote efficiency, competition, and capital formation.” Importing from antitrust law, the most plausible candidate for the overall maximand for financial market regulation is “investor welfare.” Other interests will tend naturally to feed into this maximand rather than conflict with it as transaction costs tend toward zero. In any event, if it is to move forward with any kind of cost-benefit analysis the SEC will have to abandon its fixed-pie, out-of-equilibrium version of market exchange.

SEC’s Current Guidance memo suffers from a troubling omission regarding the notice and comment process. Market participants will often hesitate to explain their resistance to a regulation in a public comment if doing so risks revealing confidential business information. Few market participants want to publicly divulge how their rivals

97 It is facile to show that market insiders generally make greater returns than small investors. In what market is this not the case? Real estate? Art? Minerals? Intellectual property? Insiders earn greater returns because they invest more and put more capital at risk. That small investors can enter the market and expect pretty good returns without being fleeced is a testament to the power of economic institutions, including private ordering.
might take advantage of them. Better to keep quiet and hope for the best. One course of action for SEC staffers is to recognize that such situations are possible and that public comments might approach an issue obliquely in a way that does not appear to make sense on its face. All the more reason to have a well-developed economic understanding of the business practices at issue before regulating. To protect proprietary information, the SEC should give serious thought to allowing some form of public comment or other input that does not require commenters to reveal the substantive reasons for favoring or opposing a given regulation. Perhaps conducting anonymous surveys or recording a simple anonymous vote of the affected market participants is in order.

The SEC’s Current Guidance memo lists as examples of market failure “. . . principal-agent problems (such as economic conflicts of interest), and asymmetric information.” SEC rulemaking has routinely identified the elimination of conflicts of interest as justification for regulation. This approach can be used to justify any regulation. Conflicts of interest are an inevitable artifact of productive specialization and should not be eliminated wholesale lest the cure be worse than the disease. Every conflict of interest creates the potential for self-dealing, which is, of course, a costly form of wealth transfer. But the parties to a category of exchange have the incentive to adopt contractual terms and other institutional arrangements to avoid costly transfers. Only transaction costs stand in their way. Identifying the existence of a conflict of interest may be a necessary condition for regulation, but it is far from sufficient. To justify a regulation because it reduces self dealing, the SEC must show that the regulation would reduce the transaction costs the parties face to limit their own inefficient attempts to capture wealth.

Much the same reasoning applies to information asymmetry as the justification for regulation. Let us return to the parable of the eggs to see why. In the real world my egg trade with Crusoe might be less than ideal because of transaction costs, but transaction costs are real costs. To increase the gains from trade, Crusoe and I might adopt various institutional arrangements to lower transaction costs. Suppose, for

---

98 The SEC has shown modest appreciation for the general point that disclosure can make a problem worse. In ICA Rel. No. 26418 (April 16, 2004) it withdrew a proposal to require funds to provide specific disclosure of how they detect market timing because the disclosure would likely provide market timers with a “road map” to evade detection.

99 See supra note 54, at 5.
example, that all eggs are not created equal, and that Crusoe fears I will strategically slip him the bad ones. Identifying the bad eggs, of course, takes my time and effort — a transaction cost. And Crusoe will take time and effort to inspect each batch, lowering the price he is willing to pay. I will react by selling him fewer eggs. Because of transaction costs, aggregate net social benefits will be less than ideal.

To counteract this threat, Crusoe and I might agree to a long course of dealing for which I post a sunk performance bond in the form of reputational capital (brandname) and then charge a premium price for the assurance it provides.\textsuperscript{100} Crusoe then recognizes that I have too much to lose by strategically slipping him the bad eggs because he would eventually terminate me and I would lose my bond and reputation. To the extent he trusts me to give him a representative sample of eggs he need not become an expert in assessing their quality.\textsuperscript{101} He is willing to pay slightly more than \(P^*\) without inspection and is better off being \textit{asymmetrically informed}. At the same time, I have nothing to gain by strategically slipping him the bad eggs, so I avoid spending the extra resources necessary to do so. Asymmetric information is endogenous — driven by specialization and prevailing institutional arrangements. It saves society’s scarce resources by avoiding information duplication. When we observe two asymmetrically informed people trading repeatedly it is precisely because they have resort to trust arrangements such as brandnames that makes the less-informed party comfortable and both parties more willing to specialize. Asymmetric information is a condition that should be encouraged, all else being equal, rather than regulated out of existence.

This does not mean informed market participants never take advantage of their less-informed trading partners in an attempt to capture wealth, only that doing so is inefficient and provides market participants with the impetus to develop institutional arrangements that avert costly wealth transfer. This explains why repeat dealing based on relationships of trust is so common in financial markets. It also explains why market insiders must be well compensated; the ongoing benefits of being trustworthy must outweigh the one-time gains from cheating if cheating is to be deterred. Through

\begin{footnotesize}
\begin{enumerate}
\item \textsuperscript{100} Klein & Leffler, \textit{The Role of Market Forces in Assuring Contractual Performance}, 89 J. Pol. Econ 615 (1981).
\end{enumerate}
\end{footnotesize}
rulemaking and other actions, regulators might be able to reduce transaction costs enough that the parties’ cost of developing trust arrangements falls even more. This is the case that must be made under TCBA.

C. Transaction Cost Economics and SEC Rulemaking

This Part addresses two issues. First, it looks at the SEC’s past practices regarding the economic justification for rulemaking. It does so primarily in the context of investment adviser regulation. I devote specific attention to the regulation of mutual funds and their advisors and the fees advisers charge their managed funds. For over 50 years the SEC has embraced an economic model in which advisers face little in the way of outside competitive pressure. It has steadfastly clung to the mistaken belief that untapped economies of scale exist in fund management. With market competition weak or absent, advisory fees are therefore excessive and fund boards ineffective. I show that this model is hopelessly flawed. This requires a close examination of the mutual fund form of organization as well as a careful look at the fund-adviser relationship. Next, this Part looks at the SEC’s recent misuse of cost-benefit analysis. It begins by reviewing the D.C. Circuit Court’s decision in Chamber I in light of TCBA. It then provides a TCBA analysis of the SEC’s 2004 rule requiring standardized disclosure of the effect ongoing advisory fees have on fund returns.

1. The SEC’s Past Practice Regulating Mutual Funds and Their Advisers

a. A Brief Introduction to Mutual Funds and Investment Advisers

Mutual funds are investment pools organized as corporations or trusts under state law. To raise capital, the fund issues shares to the investing public. The promoter places most of the proceeds in an efficiently diversified portfolio of risky assets (primarily corporate stocks and bonds, government debt, etc.) and cash to which shareholders have an undivided per share claim. The unique feature of mutual funds — often referred to as open-end funds — is that they offer to issue and redeem shares publicly at the net asset
value of the portfolio next calculated, normally at the close of trading on the exchange where the portfolio securities are most actively traded. Net asset value (NAV) is equal to the reported prices of the underlying portfolio securities less any transaction costs, advisory fees, and administrative expenses the fund incurs since the last calculation.\textsuperscript{102} Because the fund offers shareholders redemption rights, there is no need for a secondary public market in mutual fund shares. In contrast, shares in closed-end funds, once issued, must be bought and sold in secondary public markets through brokers or dealers.

Traditionally, broker-dealers sell the shares issued by mutual funds to the investing public. Early on, an underwriter affiliate of the adviser performed the distribution function through its retail broker-dealer network. As compensation for their marketing efforts, broker-dealers were paid an up-front sales load,\textsuperscript{103} normally about five percent of the purchase price of the shares. Sales loads have long been subject to minimum resale price maintenance under ICA Rule 22(c) prohibiting any member of a self-regulatory organization such as the New York Stock Exchange from charging a smaller load.\textsuperscript{104} The standard explanation for resale price maintenance from transaction cost economics clearly applies in this setting. The fund’s board, perhaps at the behest of the adviser, wants broker-dealers to compete in selling fund shares to the investing public by providing point-of-sale services rather than by cutting the load. Examples of special services a broker-dealer should provide are adhering to know-your-customer rules, helping customers select knowledgeably between alternative investments, and making customer suitability determinations.

Over time the use of loads to compensate broker-dealers has gradually declined with the ascent of so-called “no-load” funds. Rather than having shareholders pay a one-time load, no-load fund shareholders now collectively pay recurring 12b-1 fees out of the common assets of the fund.\textsuperscript{105} Rule 12b-1 has had two important effects on the market for mutual funds. First, it has dramatically reduced investor switching costs. Second, it

\textsuperscript{103} "Sales load’ means the difference between the price of a security to the public and that portion of the proceeds from its sale which is received and invested or held for investment by the issuer.” 15 U.S.C. § 80a-2(a)(35) (2006).
\textsuperscript{104} Citation. Does this require a BOD decision? Check that SEC regs prohibited cutting the load.
\textsuperscript{105} Id. Note that 12b-1 fees uncouple the sales charges from the investor who causes them to be incurred.
has allowed fund advisers to market shares through multiple unaffiliated broker–dealers by way of so-called “fund supermarkets” rather than through affiliated brokers. With broker–dealers offering shares in a large number of funds sponsored and managed by different advisers, the conflicts of interests once thought to plague the marketing of shares exclusively through affiliated broker-dealers have been reduced.

Section 15 of the ICA generally requires mutual funds to be managed by an external advisory firm pursuant to a contract approved by the fund’s outside directors approve. In exchange for advisory and other services, the contract pays the adviser a periodic percentage fee based on total assets under management, normally somewhere on the order of 10 to 150 basis points per year. By way of example, if the fee is 45 basis points and total assets under management during the entire year are $100 million, the adviser’s fee for that year is $450,000 (.0045 x $100M). Most funds are managed by a central adviser within a complex or family of funds, in which case the hands-on manager of each fund is an individual or team employed by the advisory firm or an external sub-adviser under contract.

Mutual funds can be either actively or passively managed. Passively-managed index funds place their assets in a portfolio of securities that mimics a published index such as the S&P 500. Their portfolio composition changes only when the composition of the index changes, such as in 2008 when General Motors dropped out of the list of the 500 most highly capitalized U.S. corporations based on the reduced value of its outstanding shares.

Aside from ministerial functions, in an actively-managed fund the advisory function consists most importantly of portfolio management. As an agent for the fund, an active manager’s primary charge is to hold an efficiently diversified portfolio, to use his best efforts to perform or acquire research to identify mispriced securities, and to buy or sell those securities to make a profit for the fund before the market fully corrects the

---


107. Id. at 115.

108. Id.

109. The S&P 500 index is a hypothetical portfolio whose securities consist of the common stock issued by the 500 largest U.S. corporations by total market capitalization. The weight of each stock in the index is equal to the corporation’s equity capitalization relative to the total of all 500 firms.
pricing error. In buying or selling securities for the portfolio the manager has a fiduciary duty of “best execution.”

Active fund managers are free to vary the composition of their portfolios provided they meet certain statutory diversification and regulatory disclosure requirements based on their reported “style” (large cap, small cap, growth, income, value, biotech, etc.). To facilitate shareholder monitoring of fund managers’ compliance with their stated styles, in 2004 the SEC required managers to report their portfolio holdings to the SEC quarterly with a 60-day lag and to provide standardized expense reporting.110

In addition to mandating vertical separation between the adviser and the fund, the ICA imposes various other substantive requirements on funds and their advisers.111 Even though the adviser normally creates and promotes the fund from its inception, Section 15(a) requires the fund board or shareholders to periodically renew the advisory contract and that the contract be terminable at will on 60 days written notice.112 Boards are composed of both inside directors (those affiliated with the adviser) and outside directors. The ICA mandates that at least forty percent of the board consist of outside directors, but any funds with a 12b-1 plan in place are required to have sixty percent outside directors. Section 15(a) requires that advisory contracts terminate on assignment, although section 15(f) allows advisory firm owners to profit from a sale of control in the advisory firm that results in an indirect assignment if certain conditions are met.113 Section 36(b), added in 1970, imposes on fund advisers a fiduciary duty with respect to the receipt of compensation for services and provides shareholders with a private cause of action for damages limited to the compensation received up to one year before the action was instituted.114

Given that the adviser normally creates the fund and makes long-term relationship-specific investments in promotion, it is unsurprising that advisory contracts are almost invariably renewed. It is rare for the board to shop the contract around, which in part accounts for widespread criticism that outside board members suffer from a

112 Id.
113 Id.
114 § 80a-35(b).
structural bias in favor of the adviser. The *de facto* relationship between the adviser and the fund lies somewhere in an economic netherworld between vertical integration (an extended firm) and long-term contract (market exchange), with the standard fiduciary and trustee duties being variously applied.\footnote{See generally R.H. Coase, *The Nature of the Firm*, 4 ECONOMICA 386 (1937).}

The mutual fund industry is highly competitive by any structural standard on which the FTC or DOJ would rely. There are literally thousands of mutual funds in the U.S. Industrial concentration is low and has declined over time, and both innovation and entry are prolific.\footnote{See Coates & Hubbard, supra note 119, at 163–81 (describing the recent history of competition in the mutual fund industry).} The concentration of fund families is substantially higher than for mutual funds alone, but fund family concentration is nonetheless low and declining, and no single fund or fund complex enjoys a dominant market share.\footnote{Id.}

b. **Down a Rabbit Hole — The SEC’s Mistaken View of Advisory Fees**

At the SEC’s request, in 1962 the Wharton School of Finance and Commerce at the University of Pennsylvania published the influential *Wharton Report*,\footnote{Wharton Sch. of Fin. & Commerce, *A Study of Mutual Funds*, H.R. REP. NO. 87-2274 (1962). Irwin Friend was the lead investigator.} a study of the mutual fund industry. With Irwin Friend as its principal investigator, the *Wharton Report* found that total assets under management and fund size in the industry had grown dramatically during the study period. Yet fund advisers had done little to reduce fees despite what the *Wharton Report* asserted were clearly fixed costs of management that do not vary with assets. As a result, economies of scale in fund management must be present, and fees should have declined over time owing to market competition.\footnote{John C. Coates IV & R. Glenn Hubbard, *Competition in the Mutual Fund Industry: Evidence and Implications for Policy*, 33 J. CORP. L. 151, 156 (2007).} It concluded that fee competition in the industry was weak or altogether absent.\footnote{Id.} The SEC followed in 1966 with a report to Congress drawing the same conclusion and recommending various statutory amendments to protect fund investors from excessive fees.
fees, including fee caps unless industry members policed themselves by lowering fees to account for scale economies. The industry capitulated, with many fund advisers adopting stepped-down management fees as assets under management grew. In the interim, Michael Jensen (1968) published a study showing that NAV returns to actively-managed funds fell short of the return to the S&P 500 market portfolio during his study period. He concluded that active management added nothing to shareholder returns. Congress responded in 1970 by amending the ICA to add section 36(b), imposing on fund advisers the prospect of private suits for excessive fees.

The Wharton Report and the SEC’s response to it may be one of the most unfortunate developments in the SEC’s history. The entire episode completely ignored obvious concurrent developments in industrial organization that were to prove remarkably influential in antitrust. While the Report appears thoroughly researched and carefully crafted (good reading for insomniacs), it is a mastery of the kind of shoddy industrial organization economics that plagued the early years of antitrust. By way of example, an obvious implication of the Neoclassical model of perfect competition is that if scale economies exist at the current scale of industry firms competition will drive the industry to be dominated by fewer and larger firms. This was clearly not the case at the time, nor has it been since. Despite the obvious, the Report’s legacy in the rulemaking process has persisted, until recently let us hope.

122 See SEC, 1966 REPORT, supra note 121, and accompanying text.
123 Note the effect of stepped-down fees in discussing the open-access character of funds.
125 New section 36(b) of the Investment Company Act states in relevant part that:

[T]he investment adviser of a registered investment company shall be deemed to have a fiduciary duty with respect to the receipt of compensation for services, or of payments of a material nature, paid by such registered investment company . . . . An action may be brought under this subsection by the Commission, or by a security holder of such registered investment company on behalf of such company, against such investment adviser . . . . for breach of fiduciary duty in respect of such compensation or payments paid by such registered investment company . . . . to such investment adviser or person.

(3) No award of damages shall be recoverable for any period prior to one year before the action was instituted. Any award of damages against such recipient shall be limited to the actual damages resulting from the breach of fiduciary duty and shall in no event exceed the amount of compensation or payment received from such investment company . . . . by such recipient.
In hindsight, the Wharton Report’s assumption that untapped scale economies must have been present in the industry is inexcusable. To quote Nobel Laureate George Stigler — no friend of Irwin Friend\textsuperscript{126} — “[t]he theory of the economies of scale is the theory of the relationship between the scale of use of a properly chosen combination of all productive services and the rate of output of the enterprise.”\textsuperscript{127} Scale economies exist when the average cost of producing an economic good consumers demand declines as the scale of production rises; double all inputs and if output more than doubles the average cost of production must decline and economics of scale must be present. The foundational problem with Friend’s analysis is that a mutual fund’s total assets under management is not an output fund investors demand, nor is it an accurate characterization of how fund advisers add value, that is, what they produce.\textsuperscript{128} Indeed, all else being equal existing fund shareholders are worse off as assets-under-management increase; they would prefer at any time to close the fund to outside investors.

The ill effects of the Wharton Report may have permeated the entire agency, but it has most surely permeated the SEC’s regulation of investment advisers. It has allowed the SEC staff to completely ignore the Neoclassical model of perfect competition — the foundation of modern industrial organization economics and by far the most successful and tractable model in the social sciences — and to regulate under the implausible assumption that one group of market participants can systematically exploit another group. Friend’s theory is no theory at all because it fails to hypothesize any kind of equilibrium between fund advisers, the funds they manage, and investors. Nor does it acknowledge that investors in competitive markets have the wherewithal to protect themselves in any number of ways. As a result, it does not yield testable implications capable of refutation.

No doubt fund management is subject to fixed costs that do not vary with assets under management. So what? Through some range of output average cost may indeed decline in total assets, but a firm’s rate of output, properly defined, is determined by its marginal cost, which also determines what the manager must be paid to incur these costs.

\textsuperscript{127} GEORGE STIGLER, \textit{THE ORGANIZATION OF INDUSTRY} 71 (1968).
\textsuperscript{128} This point boils down to the elementary proposition that when we draw marginal value and marginal cost on the blackboard, they must both match up with what is being measured on the horizontal axis.
Just because the average cost of management declines in fund assets is no reason to conclude that competitive asset-based advisory fees should decline in fund assets. To draw any such conclusion requires an economic theory of contract choice defining the good being transacted and explaining how, in equilibrium, adviser compensation affects the cost of transacting and the joint gains from trade.

2. The SEC’s Recent Misuse of Cost-Benefit Analysis

a. TCBA Analysis of Chamber I

The question naturally arises how TCBA analysis of the three SEC rules struck down by the D.C. Circuit Court decisions might have contributed to the rulemaking process and the ultimate substantive rationale for or against the rule. It is impossible to know the answer with any kind of precision because had TCBA been part of the rulemaking process it surely would have altered the factual record in a material way compared to the one that emerged based on the SEC’s fixed-pie, out-of-equilibrium “economic” model. In what follows, I provide a few observations on Chamber I based on TCBA.

The SEC justified its Corporate Governance Rule as a response to then-recent scandals involving market timing and other practices, which advisory firm personnel apparently approved as part of a quid pro quo in some cases. Taking this justification at face value, the SEC had never come to grips with the economics of market timing despite very good reasons to do so.

What is fund timing? Prior to the mutual fund scandals, many advisers allowed investors to exchange shares between different funds in the complex at closing NAV. Mutual fund timing occurs when a shareholder in one fund takes advantage of the free option to exchange his shares for those of another fund in the complex at stale prices. Quite literally, this option can arise because the world turns and is often referred to as time zone arbitrage. Consider an advisor who manages a fund complex and offers free

129 An alternative, and plausible, explanation is that the SEC was embarrassed by N.Y. State Attorney General Eliot Spitzer’s identification and aggressive prosecution of these practices. The SEC and its staff may simply have decided they needed to do something to save face publicly.
exchanges between, say, its U.S. and European equity index funds. Suppose each fund has ten shareholders, each holding ten shares, and that each fund has total assets of $100 when the London market closes at 11 a.m. Eastern Standard Time. Ignoring fees for simplicity, at that moment the NAV per share of each fund is $1.00, which fixes the terms of exchange into the European fund. If the New York market rises dramatically to $110 after the close of trading in London, the European fund’s closing NAV of $1.00 per share is advantageously low because all stock prices tend to move together. If the London market were open the European fund would almost invariably reflect the U.S. market movement. An aggressive investor can earn an almost riskless profit shortly before the New York close of trading by exercising his free option to exchange shares in the U.S. fund, accurately valued at $1.10 each, for shares in the European fund, under-valued at $1.00 each.

The problem with fund timing is that it dilutes the value of passive investors’ shares. Suppose each of four market timers in the complex exchanges his 10 shares in the U.S. fund for 11 newly issued shares of the European fund. When the London market opens the next morning the value of its underlying securities will immediately increase to $110 to reflect the updated valuation of its portfolio securities. In addition, it will have $44 in cash, which it received from the U.S. Fund as a result of timer exchanges. Its total asset value of $154 must now be divided equally between 144 shares, for only $1.07 per share. Absent the actions of timers, complacent investors in the European fund would have 100 shares worth $1.10 each. As a group they lose a total of $3.00 to active investors ($0.03 per share times 100 shares). Timing works in the opposite direction for market downturns, so that passive investors in either fund can suffer dilution.

Imbedded in the SEC’s Corporate Governance Rule is the assumption that market timing was just another example of insiders taking advantage of defenseless investors. Its response to Spitzer-driven public revelations of market timing would have made even Casablanca’s Inspector Renault (Claude Reines) blush after declaring he is “Shocked! Shocked to find that gambling is going on here” as his assistant hands him his winnings. The SEC had long known of market timing.\footnote{See Jerry W. Markham, Mutual Funds Scandals - Comparative Analysis of the Role of Corporate Governance in the Regulation of Collective Investments, 3 Hastings Bus. L.J. 67, 82-84 (2006); Conrad S.} On discovering market timing and its
diluting effects on passive investors, fund advisers attempted to shut the timers down by using “fair-value pricing”\(^{131}\) or by imposing exchange fees or trade limitations on them. Some timers threatened legal action on the theory that they had bought shares under a prospectus that allowed them free exchanges, and that they were therefore grandfathered in. Many complained to the SEC that funds’ use of fair-value pricing was abusive. The SEC’s response to advisers’ entreaties for guidance was that they should adhere to the language in their prospectuses, whatever that happened to be. In other words, they got no guidance. Here was a case in which the SEC could have taken action to reduce market participants’ costs of transacting but declined to do so.\(^{132}\)

The question is why advisers ever allowed intra-family exchanges given the conflict of interest market timing creates between aggressive and passive shareholders? Not surprisingly to anyone familiar with transaction cost economics, fund timing is not the only conflict advisers must address. And in equilibrium addressing multiple conflicts inevitably means balancing them. By way of example, rational investors in a mutual fund complex (or any other multiple account setting) naturally fear the advisor will secretly commingle assets by allocating investment research, brokerage expenses, overhead, and even investment performance in a way that strategically builds one fund’s returns at the expense of others. Favoritism might be tempting for the adviser because superior returns in one fund would ordinarily generate dramatic inflows from outside investors, while lackluster returns in other funds would result in fewer offsetting outflows by existing shareholders, at least in the short run. The net effect could be increased management fees to the adviser and lower returns at passive investors’ expense.

---

\(^{131}\) ICA Section 2(a)(41) states that registered investment company may calculate portfolio net asset value based on “fair value as determined in good faith by the board of directors” where market values of securities are not readily available.

\(^{132}\) Shortly after the market timing scandals broke the SEC made various amendments to its registration and prospectus disclosure requirements. ICA Rel. No. 26418 (April 16, 2004).
Ex ante, the prospect of favoritism is likely to deter investors from placing their money in mutual funds, reducing assets under management. To prevent this, fund advisers effectively bonded themselves against favoritism by allowing existing shareholders to freely transfer between funds in the complex.\textsuperscript{133} With at least some rational — and what this episode reveals to be aggressive — shareholders, internal transfers would quickly have dissipated excess returns in the favored fund, deterring asset inflows from outside the complex. Favoritism would have been a losing proposition for the adviser.

Rather than positing that the world is out of equilibrium, had the SEC relied on transaction cost economics it would have recognized that every business arrangement is subject to multiple conflicts of interest, and that it is the fund board’s function to balance — rather than to eliminate — these conflicts. It very likely would have either abandoned the \textit{Corporate Governance} rule or articulated a far more convincing rationale to justify it. Nothing in the above explanation for market timing suggests that free intra-family exchanges were inherently bad, that inside directors were playing hide-the-peanut with outside directors, or that increasing the number of outside directors or installing an outside chairman would have changed anything.

b. ICA Rule 30 — Standardized Fee Disclosure

In March 2004 the SEC adopted Rules 30b1–5 and form amendments under the SA, SEA, and ICA. Titled \textit{Shareholder Reports and Quarterly Portfolio Disclosure of Registered Management Investment Companies}, it requires mutual funds to include in their quarterly shareholder reports standardized disclosure of the effect of ongoing fund fees on portfolio returns covering the reporting period. It also requires funds to report a summary schedule of portfolio holdings as of the end of the covered quarter and to file with the SEC a detailed schedule of portfolio holdings, to be made available to shareholders free of charge on request. Quarterly reports must be filed within 60 days of the end of the quarter. The purpose of the rule and amendments is to “provide better

\textsuperscript{133} To be accurate, the fee would only have to be lower than the fee outside investors would have to pay to enter. In load funds the exchange fee could be greater than zero and still have the desired bonding effect. In fees with 12b-1 fees the exchange fee would have to be zero to accomplish the same result.
information to investors about fund costs, investments, and performance.” The following provides a TCBA analysis of the Quarterly Disclosure Rule’s standardized disclosure of the effect of ongoing fund fees on portfolio returns. The Quarterly Disclosure Rule completely fails a TCBA analysis.

As background, the SEC noted that it began requiring funds to disclose fees and other charges starting in 1988. Based on its belief that fund investors do not fully understand the implications of expenses — especially recurring management fees — on fund returns, it had considered various rules over the years to correct the problem. Among other things, it posted a Mutual Fund Cost Calculator on its Web site to allow investors to compare the cost of owning different funds. None of this seems to have worked, leaving the SEC with significant concerns regarding “the degree to which investors understand the nature and effect of these ongoing fees.” To increase investor understanding, the Quarterly Disclosure Rule requires that henceforth shareholder reports must include:

(1) the cost in dollars associated with an investment of $1,000, based on the fund’s actual expenses and return for the period; and (2) the cost in dollars associated with an investment of $1,000, based on the fund’s actual expenses for the period and an assumed return of 5 percent per year. The first figure is intended to permit investors to estimate the actual costs, in dollars, that they bore over the reporting period. The second figure is intended to provide investors with a basis for comparing the level of current period expenses of different funds.

This passage demonstrates the SEC’s adherence to the fixed-pie, out-of-equilibrium economic model. For every dollar fund shareholders pay in advisory fees, it suggests that their returns necessarily fall by a dollar. High-fee funds must be exploiting investors. The underlying policy implication is that if advisers could be convinced or compelled to reduce their fees — consistent with the SEC’s earlier threat to impose fee caps — investor returns would rise. The reasoning is peculiar on its face if advisers face

134 69 Federal Register 46 (March 9, 2004), at 1244-45.
any kind of market competition. If high-fee funds can so easily exploit investors free from competitive forces, why not double or triple fees? Alternatively, with the availability of low-fee index funds, why would investors not simply exit high-fee funds en masse? Even more peculiar is that the SEC ignored empirical scholarly literature from financial economics by Malkiel (1995), Carhart (1997), and others finding that, cross-sectionally, high-fee funds earned lower returns than low-fee funds. Instead, it relied entirely on reports from the financial press to support its conclusion that advisory fees reduce shareholder returns. One can find support for anything in the financial press. Using it to justify significant rulemaking is irresponsible when relevant peer-reviewed scholarship is at hand.

Perhaps the reason the SEC neglected the scholarly literature was that it would also have had to report on more recent and far more thorough empirical work by Russ Wermers (2000) that completely debunks the earlier findings. Based on SEC filings, he assembled an exhaustive database showing both the NAV returns to every U.S. mutual fund over the study period and the gross returns each fund manager generated based on actual stock picks, which he backed out of each fund’s periodic reported portfolio stock holdings. This allowed him to differentiate between manager stock picking returns and the returns shareholders earned based on NAV performance. The difference consisted of

137 The Rule’s footnote 29 reads as follows:

See, e.g., Mara Der Haranesian, et al., How to Fix the Mutual Funds Mess, Business Week, Sept. 22, 2003, at 106 (discussing the impact of fees on returns and arguing that it is difficult for investors to determine what they personally pay based on a fund’s expense ratio); Chuck Jaffe, In ‘Plain English,’ Disclosure is a Joke, The Boston Globe, August 31, 2003, at E4 (arguing for more understandable fee disclosure in fund prospectuses); Theo Francis, Getting the Most From Fund Costs, Wall street Journal, Dec. 2, 2002, at R1 (discussing the importance of considering fees and expenses when investing in mutual funds, and explaining how to use the SEC’s cost calculator); James Glassman, A Failing Grade for Mutual Funds, Washington Post, Dec. 1, 2002, at H1 (discussing importance of differences in expenses to fund returns, and using examples from SEC’s cost calculator); Neil Weinberg, Fund Manager Knows Best; As Corporations are Fessing Up to Investors, Mutual Funds Still Gloss Over Costs, Forbes Magazine, Oct. 14, 2002 (84% of investors believe higher expenses result in higher performance); Investors Need to Bone Up on Bonds and Costs, According to Vanguard/Money Investor Literacy Test, Press Release, Business Wire, Sept. 25, 2002 (75% of survey respondents could not accurately define fund expense ratio and 64% did not understand the impact of expenses on fund returns).

various transaction costs, fees, and administrative expenses.\textsuperscript{139} He found that active funds outperformed the market by 1.3% per year but that their NAV returns (the returns shareholders enjoyed) fell short of the market by 1%. He accounted for 0.7% of the 2.3% difference as the result of underperforming non-stock holdings such as cash, necessary to meet share redemptions.

The remaining 1.6% difference consisted of expenses, including advisory fees, and transaction costs. One of the problems a skillful manager faces is that to make a profit he must incur transaction costs to trade the securities he believes are mispriced. Transaction costs, which accrue to the fund, consist of brokerage commissions and any adverse price impact on trades between the moment the manager orders the trade and the moment the broker fully executes it. Price impact results from the mere presence of the manager in the trading environment. If the trade is large and the manager is known for being informed, this information can leak out, causing the price of the security to move against him before he can complete it. Price impact can be a substantial drag on portfolio performance. The observation that some managers’ trades consistently experience large price impact suggests they represent “smart money” in the securities trading environment. Wermers found that although high-turnover funds (presumably a proxy for smart money) had higher expenses and transaction costs, their net stock picking returns exceeded those of both low-turnover funds and the Vanguard 500 Index fund, thus casting serious empirical doubt on Malkiel’s and Carhart’s findings based on NAV returns that expenses and performance are negatively related.

Even as the SEC was formulating the \textit{Quarterly Disclosure Rule}, Berk & Green (2004) put forth important theoretical work showing that, if investors are rational and capital markets are efficient, fees and NAV returns should be unrelated. Even more important, fees should be irrelevant to fund shareholder returns. In their words,

\begin{quote}
[i]n the face of this evidence many researchers have concluded that a consistent explanation of [advisory fees] is impossible without appealing to behavioral arguments that depend on irrationality or to elaborate
\end{quote}

\textsuperscript{139} He later likened the ability to assess manager stock-picking returns based on portfolio holdings, rather than NAV returns, as akin to advances in DNA profiling over fingerprint methods in criminal proceedings. \textit{Id.} at 208.
theories based on asymmetric information or moral hazard. One thing that has been missing from this debate is a clear delineation of what a rational model, with no moral hazard or asymmetric information, implies about flows and performance. Before appealing to these additional effects, we believe that it makes sense to first establish which behaviors in the data are qualitatively and quantitatively consistent with more direct explanations.\(^\text{140}\)

In their model, fund size and total manager compensation increase in response to perceived manager skill until returns are equalized across funds going forward. Managers’ failure to outperform a passive benchmark based on NAV returns does not mean they lack skill. “It merely implies that the provision of capital by investors to the mutual fund industry is competitive. . . . [Owing to fund flows] investors cannot expect to make positive excess returns, so superior performance cannot be predictable.”\(^\text{141}\)

In the language of transaction cost economics, from investors’ standpoint a mutual fund is an “open-access commons.” As with an open-access fishery or oil reservoir, entry by investors will continue until all excess returns are competed down to the return on the next best alternative investment. The size of the fee advisers charge is irrelevant. Holding investors’ expectations of manager stock-picking skills and other factors constant between two funds, the fund with the lower fee will simply attract larger inflows in the process of equalizing investors’ expected returns with their best outside opportunity.

Figure 7 provides a simplified illustration of the competitive effect of fund flows on shareholder returns as management fees vary. The horizontal axis measures total dollars invested and the number of shares issued, \(S\), with each share priced at one dollar, as in a money market fund. The manager knows how to generate an absolute return, \(R\) (say, \$100), at the end of the year.\(^\text{142}\) The question rational investors face is how much


\(^{141}\) Id. at 1271.

\(^{142}\) The assumption of a given dollar return is somewhat unusual but also intuitive. Any MBA learns to undertake all positive net present value projects, as expressed in present-value dollars after appropriate discounting. Using the internal rate of return expressed as an interest rate can be misleading. See Brealey
money to contribute to the fund at the beginning of the year knowing they will have to pay a fee on both $R$ and their invested assets, $S$. As long as entry is open they know they must share $R$ in common. Under these conditions, investors continue to buy shares in the fund as long as the return they earn minus the fee they pay on that return, $R(1-f)$, just equals the fee they pay on their invested capital, $S$, plus the outside return they could have earned in the index fund, $r$. With a fee of $f_1$ they invest $S_1$ dollars, where $R(1-f_1) = r + f_1$. Total returns to all investors are $R(1-f_1) \times S_1$, equal to the large shaded rectangle. Total fees on the money they invest are $f_1 \times S_1$, equal to the upper shaded rectangle, and the total returns they earn are $r \times S_1$, equal to the lower shaded rectangle. As in any competitive model with open entry and exit, fund shareholders just break even, earning a return of $r$ per share.\textsuperscript{143} I investors compete for returns, which they surely do, they can never expect to earn excess returns. What is more, many of the assets invested in the fund may be unnecessary to provide the manager with investment capital. Rather, they simply serve as place-holders for shareholders’ proportionate claim to investment returns.\textsuperscript{144}

This is true even as the advisory fee varies. Suppose the SEC imposes a fee cap of $f_2$. This opens up the prospect that shareholders can earn an abnormal return. Existing fund shareholders would prefer to see the fund closed to new inflows, but the fund stands ready to issue shares to all comers at NAV, momentarily equal to $r$. At $S_1$ investors will happily pay $r$ for a share worth $R(1-f_1)$. Competition will lead investors to crowd into the fund until they have invested $S_2$, such that $R(1-f_2) = r + f_2$. Once again fund shareholders just break even, with their return per share just equal to $r$. Jensen’s (1968) findings that actively managed funds earn no excess returns, though alarming at the time, are exactly what we would expect given the open-access nature of mutual funds.

In this simple but telling model, advisory fees have no effect on shareholder

\textsuperscript{143} One implication of open-access in mutual funds is that the stepped down fee structure many funds adopted in the late 1960s under the SEC’s threat of fee caps is that these funds are now simply larger that they otherwise would be. Stepped down fees could be a barrier to entry by new funds.

\textsuperscript{144} Habib & Johnsen, Moral Hazard

& Myers, Principals of Corporate Finance (6\textsuperscript{th} ed. 2000). Beyond that, the assumption of a given dollar return captures the notion of diminishing marginal product. In this case the manager performs research to identify a single opportunity, and that is it. In reality, the more a manager invests doing research the more opportunities he or she will uncover, but at some point the value of these opportunities must diminish at the margin.
returns, and it explains why shareholders seem to pay little attention to them. Far from investors failing to understand the effect of fees on returns, it is the SEC that has failed to understand. Over the course of more than 70 years of regulating mutual funds its failure to recognize the irrelevance of fees borders on institutional incompetence. The implication of this model for cost-benefit analysis of the standardized fee disclosure in the Quarterly Disclosure Rule is that it cannot be justified. Standardized fee disclosure is costly, not just because it enlarges and complicates quarterly reports but also because it misleads investors about the effect of fees on fund returns. And because fees are irrelevant investors receive no benefits from the disclosure.

A transaction cost refinement of the above model provides further insight into the role of advisory fees and accounts for one of the apparently incriminating data points suggesting fund shareholders routinely pay excessive fees. It is that the fee per dollar of assets retail investors pay is substantially higher than what institutional investors pay, sometimes to the same manager. Far from being incriminating, this evidence is exactly what a transaction cost model of fund advisory services predicts.

Active fund management is the prototype of an experience good subject to moral hazard. Even if investors know their manager’s inherent portfolio selection skill, it takes time for them to determine whether he has spent costly effort researching portfolio selection. A manager might promise to incur costly research effort in exchange for a fee equal to marginal cost and then cheat by closet indexing to avoid these costs. To the extent investors can be temporarily fooled in this way, the manager stands to earn a one-time surplus from shirking. Knowing this, investors refuse to pay a fee that covers the manager’s research cost. The result is a low-quality equilibrium. The solution is to pay the manager a quality-assuring premium, or efficiency wage, in excess of his marginal research cost that could persist indefinitely. Because the manager’s

---

145 For a TCBA analysis of the rules requirement that funds provide complete quarterly disclosure of portfolio securities, see D. Bruce Johnsen, The Limits of Mandatory Disclosure: Regulatory Taking under the Investment Company Act (unpublished working paper).
compensation for performance in any period is paid out over time, the per period fee can be much lower than a one-time fee while ensuring that fund assets meet or exceed what the manager can profitably invest.149

This model fits the form of manager compensation in mutual funds remarkably well. What few scholars have recognized is that standard asset-based management fees are recurring. Assuming a 50 basis point annual fee, a manager who increases total assets through investment performance by $100 can expect to earn 50 cents this year, 50 cents the next year, and so on, as long as the wealth increase persists. Management fees are therefore back-end loaded but conditional on continuing satisfactory performance. Either board termination or shareholder withdrawal can truncate the stream. To the extent reputational markets work, the threat of termination deters the manager from shirking if the capitalized value of the premium fee exceeds the one-time gain from shirking. Investors are assured they will earn the normal return they bargained for and thereby benefit from higher fees.

The theory on which this model relies comes from well-regarded, mainstream scholarship from transaction cost economics.150 In mutual fund management, institutional investors pay lower fees than retail shareholders precisely because they can and do spend substantial resources to monitor the quality of the manager’s performance. Individually, retail investors lack the wherewithal to immediately and directly assess manager quality, and in any event they face a collective action problem doing so. Collectively, they can and do monitor to some extent, but by paying active managers a premium, back-end loaded fee they need not do so as aggressively as they might. Since higher fees do not come at the expense of returns, they are happy to save on information

149 Michel A. Habib and D. Bruce Johnsen, Moral Hazard in Mutual Fund Management: The Quality-Assuring Role of Fees (unpublished manuscript, George Mason University School of Law Working Papers in Law & Economics).

Among other things, this literature explains why parents who buy low-priced generic aspirin for themselves nevertheless buy expensive branded aspirin for their children. For their children they are happy to pay a quality-assuring premium that takes away the manufacturer’s incentive to cheat by strategically lowering quality to save on production costs. They also explain why Henry Ford paid his workers double the market wage. Being unable to assess the performance of assembly-line workers on the spot, he wanted them to face a severe threat from termination if caught performing negligently, which could easily shut down an entire assembly line.
costs and to be asymmetrically informed.\footnote{Habib & Johnsen show that, whatever the fee, the manager will invariably earn the same total compensation, which in the language of economics is a Ricardian rent accruing to his scarce stock-picking skill.} Social welfare increases because managers, who are members of society, earn a return adequate to compensate them for their socially productive effort identifying mispriced securities and investors are assured that they will not be cheated and experience sub-par returns.

This model shows that fees are indeed relevant to investor returns. But it shows that investors benefit from higher fees where manager quality is unobservable. Rather than spending substantial resources monitoring manager quality, shareholders pay premium fees to bond the manager to his promise. In part, this is reflected in the brandname of the fund or fund family. This allows them to be asymmetrically informed. The \textit{Quarterly Disclosure Rule}'s standardized fee disclosure is unjustified. It is costly but provides no benefits to investors beyond what they enjoy from traditional fee disclosure going back to 1988.

V. Summary and Concluding Remarks

Unfortunately, the preceding discussion is purely by way of example. There are many SEC rules that could be subjected to TCBA with equally unsettling conclusions about their effects on social welfare. Many have proven damaging to market participants, including small investors.

Consistent with the language of the statutes, and following Robert Bork’s template, the SEC must identify and follow an overall maximand to balance the offsetting effects of a regulation on the public interest, investor protection, and the promotion of competition, efficiency, and capital formation. The SEC’s \textit{Current Guidance} memo may have taken a step in this direction by suggesting that these effects be integrated into a single section. In terms of how to integrate TCBA into the internal fabric of the rulemaking process, the SEC can look to informative templates developed at the FTC and DOJ and adapt them to its particular circumstances. Following Bork’s template, it should
seriously consider the simple proxy that a proposed rule is justified on transaction cost grounds if it is likely to increase “output” of the good being transacted.

Transaction cost-benefit analysis may provide little help in a large number of settings in which transaction costs are high, but it is a tractable and intellectually rigorous framework within which to assess corrective regulation in modest-transaction-cost settings. The informational burden it imposes on regulators is substantially less than traditional cost-benefit analysis in many cases. Transaction cost-benefit analysis stands to provide a barrier to rash, misguided, or ill-motivated regulation. A substantial body of careful theoretical and empirical scholarship exists to guide financial regulators, just as it has guided regulators and federal courts in the field of antitrust for decades. No doubt in many cases a plausible application of TCBA to financial markets will lead to the conclusion that there is nothing the regulator can do to reduce the parties’ cost of transacting and that any proposed regulation is impossible to justify.

The tendency at the Division of RSFI has been to think that because the SEC regulates financial markets it should hire primarily financial economists. In general, SEC financial economists have shown excellent analytical and empirical skills, but many of them have had little training in traditional industrial organization or transaction cost economics. Given what they do, transaction cost economics is just not on their radar screen, nor should it be unless their policy recommendations warrant otherwise. But much of the business activity that occurs in financial markets is of the garden-variety kind antitrust regulators have routinely addressed using industrial organization and transaction cost economics. Going forward, the Division of RSFI should seek to include on its staff a larger number of economists trained in traditional industrial organization and transaction cost economics.152

152 A question naturally arises whether SEC economists should rely on more recently developed fields such as game theory, behavioral economics, or behavioral finance. In my opinion it should resist the temptation to do so until these fields mature to the point at which they are capable of generating testable implications, have been widely tested, and have failed to meet with widespread refutation. The U.S. Supreme Court framed the relevant standard for the reliability of scientific expert testimony in Daubert v. Merrill Dow Pharmaceuticals, 509 U.S. 579 (1993). It involves the following four-part test: the body of knowledge on which an expert proffers an opinion must 1) have been subject to peer review and publication, 2) be generally accepted as reliable in the relevant scientific community, 3) be falsifiable and have withstood attempts at falsification, and 4) have a known or knowable rate of error.
Financial economics, traditional industrial organization economics, and transaction cost economics are based on theory that is testable, has been tested, and has gone largely unfalsified. The SEC should resist looking to subfields of economic theory such as game theory and behavioral economics that are either incapable of being tested or that posit out-of-equilibrium hypotheses. Doing so could send the SEC down another rabbit hole.

In many settings, the information burden of TCBA will be substantial. In that event, the virtue of TCBA lies in the framework it provides regulators and market participants to engage in a meaningful conversation that has the potential to iterate toward first-best solutions over time as market participants innovate transaction cost reducing institutional arrangements. The goal for those charged with regulating in a modest-transaction-cost setting should be to eventually understand the cost of transacting in a given regulatory setting sufficiently well that some system of property rights can emerge as a viable alternative to regulation. In that case the parties can be left to fend for themselves with the aid of standard private law rules. In addition to being helpful to regulators sincerely focused on promoting investor welfare, TCBA offers a viable constraint on regulatory capture and special interest rent seeking. This is because it is far more tractable than traditional cost-benefit analysis in the modest-transaction-cost setting. It limits the regulator’s inquiry to identifying the relevant parties, the good they hope to transact, the economic basis for their interaction, and the transaction costs that inhibit their exchange. It also mobilizes the information and incentives of the men and women on the spot.
Figure 1

*The Neoclassical Model*
Figure 2
Tax Incidence
Figure 3
Externalities
Figure 4

How Trade Eliminates Market Failure
Figure 5

Transaction Costs
Figure 6

Double Marginalization
Figure 7
Fees versus Returns