The B.E. Journal of Economic Analysis & Policy

Contributions

Volume 9, *Issue* 1 2009 *Article* 24

Prosecutorial Discretion in Mutual Fund Settlement Negotiations, 2003-7

Eric W. Zitzewitz*

*Dartmouth College, eric.zitzewitz@dartmouth.edu

Recommended Citation
Eric W. Zitzewitz (2009) "Prosecutorial Discretion in Mutual Fund Settlement Negotiations, 2003-7," *The B.E. Journal of Economic Analysis & Policy*: Vol. 9: Iss. 1 (Contributions), Article 24.
Available at: http://www.bepress.com/bejeap/vol9/iss1/art24

Prosecutorial Discretion in Mutual Fund Settlement Negotiations, 2003-7*

Eric W. Zitzewitz

Abstract

This paper examines the negotiated settlements of 20 market timing and late trading cases, comparing the restitution obtained for shareholders with an estimate of shareholder dilution. This restitution ratio varies from 0.04 to 5, or from 0.1 to 10 if penalties are included. While some of this variation is explained by differences in the defendants' conduct, controlling for this, settlement negotiations that involved New York as well as the Security and Exchange Commission (SEC) resulted in restitution ratios that were higher by a factor of 5-10. An analysis that uses the firms' headquarters location and customers' state of residence as instruments for New York's involvement suggests that this difference is causal, and not the result of New York involving itself in cases likely to lead to large settlements. Given the much larger staff and institutional expertise of the SEC, it is likely that these differences in outcomes are due to differences in effective aggressiveness, not prosecutorial resources. Differences in aggressiveness are consistent with popular conceptions of the regulators' career concerns, as well as with theories of industry focus and regulatory capture.

KEYWORDS: Security and Exchange Commission, regulatory capture, administrative law, government accountability office, market timing, late trading

---

*Associate Professor of Economics, Dartmouth College, 6016 Rockefeller Hall, Hanover NH 03755. Tel: (603) 646-2891. email: eric.zitzewitz@dartmouth.edu. The author would like to thank Cindy Alexander, Kenneth Arrow, Severin Borenstein, Jeremy Bulow, Elizabeth Casscio, John Chalmers, John Coffee, Stefano DellaVigna, David Dubofsky, James Freyer, Robert Hall, Eric Helland, Michael Herron, Michael Kremer, Fiona Scott Morton, Jon Reuter, Nancy Rose, Richard Ruback, Bruce Sacerdote, Richard Schmalensee, Carl Shapiro, Christopher Snyder, Steven Tadelis, Peter Tufano, Matthew White, Justin Wolfers, and seminar participants at Berkeley, Dartmouth, Stanford, the 2008 American Law and Economics Association meetings, the 2008 Claremont Securities Fraud Ligitation conference, the 2009 ASSA meetings, and the 2008 IDC-Caesarea Center Finance conference for helpful discussions and comments. I also thank Charles Biderman of TrimTabs and Andrew Clark of Lipper for sharing data and for helpful conversations about their data. Over the last nine years, I have consulted, usually on a short-term basis, for many types of entities with an interest in these cases, including state and federal regulators, mutual fund firms, plaintiffs' and defense attorneys, litigation consulting and accounting firms, and a fair value pricing service, and I have benefited from conversations with many anonymous individuals.

I. Introduction

In the last decade, the Security and Exchange Commission (SEC) has faced criticism that it has been excessively passive in addressing systematic problems such as conflicts of interest faced by analysts, asset managers, and rating agencies.[1] Some have attributed this to capture by the industries it regulates, with a so-called "revolving door" staffing model for SEC attorneys cited as a root cause.[2] Others though have argued that a desire to mitigate this impression of capture has actually led the SEC to become excessively aggressive in recent years (Macey, 2005), sometimes citing increases in the dollar value of penalties as evidence (Table 1). Ultimately, these claims are difficult for outsiders to evaluate since in most cases one is uncertain about what a regulator could have achieved given the evidence and resources available.

This paper studies a rare set of regulatory cases for which these problems are partially surmountable: twenty negotiated settlements of mutual fund market timing and late trading cases from 2003-7. Unlikely most other securities regulation cases, in these cases the conduct at issue is relatively similar across cases and the harm to shareholders can be estimated using publicly available information. In particular, I calculate a restitution-to-harm ratio and restitution-plus-penalties-to-harm ratios for each settlement. Given the prosecutors stated objective of making affected shareholders whole, one would expect the restitution ratio be at least one. In order to deter future violations, the restitution-plus-penalty ratio should be at least *1/p*, where *p* is the *ex ante* probability that the conduct would be detected within the five-year statute of limitations (Becker, 1968).

[1] Examples include Pender, Kathleen, "Wherefore Art Thou, SEC?" *San Francisco Chronicle*, 12/14/2003, I1 and Dwyer, Paula "Breach of Trust," *Business Week*, 12/15/2003, 98.

[2] See, for example John Freeman's (2004) Congressional Testimony, p. 1: "The SEC is MIA. … The SEC's division of investment management (DIM) presents a classic case of 'regulatory capture'. … And then there are the DIM alums. … My analysis of SEC personnel movements, using data I obtained from the SEC under FOIA, shows that most of the SEC's senior personnel who leave the DIM go to work for mutual funds as officers or directors, for the Investment Company Institute, or for service suppliers (law firms or accounting firms). … When I was working with the SEC in DIM years ago, I was told by a fellow staff lawyer: 'Let's face it, in five years we'll all be working for these guys.'" See also Woodward (2001).

Table 1. SEC enforcement cases and recoveries by year

	Dollars obtained in judicial and administrative orders ($ millions)			
Year	Restitution/Disgorgement	Penalties	Total	Enforcement actions
1990	589	13	602	304
1991	119	11	130	320
1992	558	221	779	394
1993	225	29	254	416
1994	730	34	764	487
1995	994	34	1,028	486
1996	325	67	392	453
1997	214	49	263	489
1998	426	51	477	477
1999	650	191	841	525
2000	445	43	488	503
2001	478	44	522	484
2002	1,293	101	1,394	598
2003	900	1,100	2,000	679
2004	1,900	1,200	3,100	640
2005	1,600	1,500	3,100	630
2006	2,300	975	3,275	574

Source: SEC Annual Reports

The actual restitution ratios vary from 0.04 to 5, or from 0.1 to 10 including penalties. Proxies for ability to pay and the egregiousness of conduct explain only a small portion of the considerable variation in this ratio. In contrast, whether the New York Attorney General participated in the negotiations is strongly correlated with the restitution ratio. Of the 20 SEC-negotiated settlements in my sample, only four did not involve New York, but these four settlements included the first, second, and third-lowest restitution ratios.[3] The dilution-weighted average restitution ratio is 0.07 for settlements that did not involve New York and 0.78 for the settlements that did.

This correlation between the restitution ratio and New York's involvement, which persists when controlling for observable case characteristics, suggests two possible causal mechanisms: 1) New York's involvement caused larger settlements or 2) New York's probability of being involved in a case is correlated with unobserved factors leading to larger settlements. To attempt to distinguish between these mechanisms, I conduct an instrumental variables analysis using factors that affect New York's likelihood of involvement but are plausibly uncorrelated with a case's unobservable characteristics: whether a firm is headquartered in New York and the share of its single-state-municipal-bond-fund investors who are New York residents. The IV regressions yield estimates of a New York effect that are statistically significant and similar in magnitude to OLS estimates. This is consistent with a causal effect and inconsistent with a selection bias.[4]

Why might New York's involvement lead to bigger settlement amounts? Among state regulators, New York does have one of the tougher state "Blue Sky" anti-securities fraud laws (the 1921 Martin Act; see, e.g. Macey, 2005) and its Investor Protection Bureau is likely amongst the most sophisticated state regulators. But these cases all involved the SEC, which had both federal securities laws and its Enforcement Division (which has about 1,200 FTEs compared with 30 FTEs for New York's bureau) at its disposal. New York's involvement is not likely to be contributing significant additional manpower. [5] Some have argued

[3] The odds of 4 randomly selected settlements including the three lowest ratios are 1 in 285.

[4] There is a substantial empirical criminology literature examining prosecutor and defendant race and gender in criminal cases (e.g., Paternoster, 1984; Radelet and Pierce, 1985; Schmidt and Steury, 1989); these studies are generally forced to assume that prosecutor identity is uncorrelated with unobservable case characteristics. More recently, Helland (2001) compares Environment Protection Agency litigation outcomes before, during, and after the tenure of an purportedly pro-industry Reagan appointee, and Kondo (2006) studies the selection of arbitrators and outcomes of brokerage arbitration cases.

[5] Of course, the cases in this paper may have accounted for different shares of SEC and NYAG staff, so there need not have been a 40x difference in resources devoted to these cases. That said, it was uncontroversial among participants that I spoke with that the SEC had significantly more manpower devoted to these cases.

that the Martin Act is tougher than federal securities law in that it does not necessarily require scienter (intent to defraud) for criminal or civil actions (e.g., Dechert, 2004). In this setting, this would imply that New York could have brought cases against mutual funds that were unaware that arbitrage trading was diluting their funds. In practice, however, these cases would have been very difficult to bring for political reasons, and I am not aware that New York even threatened to bring them. The cases in this paper all involved firms that were alleged to have been not only aware of arbitrage trading, but either directly or indirectly profiting from it. Therefore in practice it seems unlikely that the no-scienter aspects of the Martin Act contributed to the toughness of laws available to regulators in these cases.[6]

A remaining possibility is New York cases obtained higher settlements because the popular impression that New York pursues these cases more effectively and/or aggressively than the SEC is correct.[7] A difference in effective aggressiveness could reflect under-aggressiveness on the part of the SEC, over-aggressiveness on the part of New York, or both. The fact that the average restitution ratio is less than 0.1 when the SEC negotiates without New York suggests the former. Restitution ratios of 2-5 for some of the cases involving New York suggest the latter.

This paper builds upon a recent report by the Government Accountability Office (2005). Cognizant of the newly large stakes involved in SEC settlement negotiations, in 2004 the Judiciary Committee of the U.S. House of Representatives commissioned a GAO review of the SEC's restitution and penalty setting process. The GAO concluded that "after the mutual fund trading abuses were uncovered in September 2003, SEC acted swiftly to bring enforcement actions against prominent firms and individuals involved in the misconduct and obtained some of its highest penalties in history from them in settlements. SEC has also consistently applied its procedures for establishing such penalties" (p. 32). At the same time, the GAO found that "SEC has not established controls that could help ensure the independence of staff from the fund industry as they carry

[6] The fact that New York's state law was tougher than other states' may have given New York more ability than other states to insist on penalties that were tougher than the SEC preferred. But this only would have been relevant if the preferences of the state and federal regulators diverged.

[7] One can find multiple examples of this view within the pages of just one major newspaper. Examples from news articles include Holstein, William J. and Edward M. Kopko, "Spitzer's Climate of Fear," *Wall Street Journal*, 11/23/2004, B2; Davis, Ann, "To Some, Santa Has a New Name: Spitzer --- New York Attorney General Turns Settlement Funds Into Gifts; Will It Grease an Election Sleigh?" *Wall Street Journal*, 12/24/2003, C1; Langley, Monica, "The Enforcer: As His Ambitions Expand, Spitzer Draws More Controversy," *Wall Street Journal*, 12/11/2003, A1. Editorials include "Mr. Spitzer's Allies," *Wall Street Journal*, 11/15/2004, A22; "Spitzer's Grandstand," *Wall Street Journal*, 3/5/2004, A14; "Spitzer's Fee," *Wall Street Journal*, 1/5/2004, A14; "Wall Street's Chaperone," *Wall Street Journal*, 4/29/2003, A16.

out SEC's critical oversight work," noting that they did not monitor whether departing staffs' future employment created a conflict of interest.

In June 2005, a Congressional hearing discussed the GAO report. In prepared testimony, it was noted that "while the GAO compares the size of the penalties to those in previous enforcement actions, an additional relevant comparison would be to the magnitude of shareholder dilution in each case" and that making this comparison would "inform the public's evaluation of the fairness of the settlement process" (Zitzewitz, 2005a, 6). In post-hearing questioning, the SEC's Division of Enforcement was asked to react to this statement. The SEC replied, in part:

> In constructing a fair and appropriate settlement package, the securities laws afford the Commission the necessary flexibility to take a case-by-vase view of the facts and circumstances of each case. … In the market timing cases, the Commission considers shareholder dilution, in itself one measure of shareholder harm, a factor in the consideration of disgorgement. … The Commission tends to refrain from identifying a single dilution figure in the public filings relating to the settlement of market timing cases, but rather focuses on the totality of the settlement package. …
>
> Consideration of the magnitude of shareholder dilution allows for a wide range of estimates of dilution to fund shareholders by abusive trades which may be available for the Commission's consideration at the time that a settlement is negotiated. This range is attributable to the many variables that may have an effect on the dollar value of particular estimates of dilution. Variables include the overall method of estimation employed, the time period used to encompass abusive trading that is subject to analysis, and the specific trades during that period that are analyzed.[8]

While declining to provide the requested degree of transparency, the SEC quote is useful in summarizing the reasons for which estimates of shareholder harm, and thus of appropriate restitution, may differ depending on choices made when constructing them. The SEC highlights three choices: the estimation method, the time period considered, and the share of arbitrage trades within a time period that the fund company was held responsible for. As I discuss in more detail below, the difference in settlements when New York was involved can be partly explained by different choices along these three dimensions.

The next section of the paper provides some background on the mutual fund cases. The third section discusses methodologies for measuring shareholder dilution, while the fourth section discusses the data I use, and my method for estimating dilution using publicly available data. The fifth section calculates

[8] U.S. House of Representatives, *Mutual Fund Trading Abuses: Hearing Before the Subcommittee on Commercial and Administrative Law of the Committee on the Judiciary*, p. 93 (available at http://judiciary.house.gov/media/pdfs/printers/109th/21653.pdf).

restitution ratios and examines their empirical determinants. The sixth section discusses possible reasons New York's involvement might have affected settlement amounts. A discussion follows.

II. Background

The market timing and late trading cases analyzed in this paper involve instances in which mutual fund management companies allowed favored shareholders to conduct arbitrage trading at the expense of their shareholders in exchange for some economic benefit.[9] Arbitrage trading in open-end mutual funds exploits the fact that fund trades are priced using an accounting calculation, rather than a market-determined equilibrium price. Essentially all U.S.-based funds value their assets once per day at 4 PM Eastern Time and use these values to calculate a net asset value (NAV) per share. This NAV is used as the price for all purchase and redemption orders placed earlier in the day. Arbitrage trading is possible when the NAV calculation does not reflect information available to the arbitrageur at the time the trade is placed.

Market timing and late trading differ in the source of arbitrageurs' advantage, as well as in their legality. Late trading cases involved trading decisions that were made after the 4 PM deadline but still received the current day's price, in violation of the forward pricing rule (rule 22c-1, promulgated by the SEC in 1968). Late traders could earn abnormal returns by simply conditioning their trading on post-4 PM market movements. In many cases, late trading was conducted without the knowledge of the mutual fund company, often by brokerages that represented to fund companies that trading decisions had been made by investors before 4 PM, when in fact some investors had been given the opportunity to cancel orders after 4 PM.[10] In other cases though, the fund management company was aware that investors were being allowed to late trade.[11] These latter cases were regarded as particularly egregious by regulators, and I find that they involved higher restitution and penalty ratios.

---

[9] The overview in this section is brief and focused on regulatory cases, as much has already been written on market timing (also known as stale price arbitrage) and late trading. Bhragava, Bose, and Dubofsky (1998), Chalmers, Edelen, and Kadlec (2001), Goetzmann, Ivkovic, and Rouwenhorst (2001), Boudoukh, et. al. (2002), Greene and Hodges (2002), and Zitzewitz (2003) discuss market timing; Zitzewitz (2006) discusses late trading. Coffee (2003) and Mahoney (2004) place the issues within the broader context of manager-investor conflicts of interest.

[10] See, for example, the cases against Security Trust Company (available at http://www.sec.gov/litigation/complaints/comp18479.pdf) or Bear Stearns (http://www.sec.gov/litigation/admin/33-8668.pdf).

[11] In November, SEC Enforcement Director Steven Cutler (2003, 16) testified before Congress that a SEC questionnaire had revealed that over 10 percent of 88 large fund families had been aware of late trading in their funds.

Market timing (also known as stale price arbitrage) trades are made before 4 PM, but exploit the fact that many fund companies use NAV calculation methods that misvalue fund shares by under reflecting recent market movements. NAVs can under-reflect recent market movements for two reasons: time zones and illiquidity. International mutual funds usually hold assets that trade on exchanges located in other time zones that close before 4 PM ET. Most mutual funds have historically valued these assets using these foreign closing prices, which do not reflect global market movements between the foreign close and 4 PM. In addition, securities that trade illiquidly often have last trade prices or bid and ask quotes that do not reflect recent market movements (Lo and MacKinlay, 1990; Kadlec and Patterson, 1999; Blume and Keim, 2006). This causes the NAVs of fund in asset classes such as small-cap equities and high-yield and municipal bonds to lag market movements. Zitzewitz (2003, Table 4) reports that in 2001, time zone arbitrage in international and global funds accounted for about 90 percent of the adverse impact on fund shareholders, with liquidity arbitrage in small and midcap domestic equity funds accounting for most of the remainder.

Unlike late trading, pre-4 PM trading to exploit inefficiencies in mutual fund valuation is not illegal *per se*. The legal issues for mutual fund companies arise from the fact that they both chose to price funds using methods that misvalued fund shares and simultaneously earned private benefits from allowing certain arbitrageurs to exploit these misvaluations. As Zitzewitz (2003) discusses, the mutual fund industry has been aware of time zone-related valuation issues for at least 25 years. Its approach, initially sanctioned by the SEC in the form of a 1981 no-action letter to Putnam Investments, has historically been to calculate fund NAVs using foreign local closing prices on most days, but to update those prices through a process known as fair value pricing on days with especially large post-close market movements. In December 1999 and again in April 2001, however, the SEC recommended far more frequent use of fair value pricing.[12] Despite this, fair value pricing was still used only infrequently by most firms before the regulatory investigations began in mid-2003 (Zitzewitz 2003 and 2005b).

Regulatory investigations later revealed that many mutual fund firms had benefited from the inefficient pricing of their fund shares by essentially selling the right to conduct large-scale arbitrage trading to favored investors. Favored investors usually paid for the right to arbitrage trade by investing so-called "sticky assets" in high-fee money market or hedge funds run by the mutual fund manager. Some of the mutual funds involved had trading fees from which favored investors

---

[12] See SEC letters to the Investment Company Institute, available at http://www.sec.gov/divisions/investment/guidance/tyle120899.htm and http://www.sec.gov/divisions/investment/guidance/tyle043001.htm, respectively.

were exempted, while others represented to investors in their prospectus that they would limit all investors to a certain number of trades per year and then selectively did not enforce this limit. In three instances, either the chairman of the fund management company or a group of portfolio managers allegedly engaged in arbitrage trading of their own funds. Like the cases in which late trading was knowingly allowed, this was regarded as particularly egregious by the regulators, and I find that these cases likewise involved higher restitution and penalty ratios.

As a means of diverting fund shareholder assets for the private benefit of the management company, misvaluing funds and then selling the right to conduct arbitrage was inefficient for two reasons. The first was that the management company was usually not able to extract the full value of the arbitrage opportunity from a paying arbitrageur. The second was that misvaluing fund shares created arbitrage opportunities for investors who did not pay for them, since fund companies could not completely exclude non-paying arbitrageurs. A crucial issue in calculating restitution is whether a fund company that misvalued fund shares and sold the right to conduct arbitrage is responsible for all of the economic harm arising from the misvaluation or for only for the dilution caused by shareholders who paid. When the SEC refers above to "the specific trades analyzed" as an important determinant of a dilution calculation, it is in large part referring to this choice.

III. Measuring Dilution

As the SEC quote above alludes to, the "method of estimation employed" can significantly affect the measurement of the dilution of long-term shareholders due to arbitrage trading. While I will use multiple methods in the analysis and while my conclusions about the effect of New York's involvement are robust to the choice of method, the method chosen does affect conclusions about the overall level of reimbursement ratios, and discussing how and why the methods differ gives some insight into the scope for prosecutorial discretion in these cases.

Broadly, there are two methods for measuring dilution. The first was developed in the pre-scandal academic literature on stale price arbitrage and late trading, which I will refer to as the "academic method." There are two versions of the academic method, but they yield highly correlated results in most samples. The first version, used by Goetzmann, Ivkovic, and Rouwenhorst (2001) (GIR) and Zitzewitz (2003), measures dilution resulting from traders purchasing (selling) fund shares for less (more) than their current value. This is given by:

$$\sum_t \Delta shares_t \cdot (NAV_t^{FV} - NAV_t^{Actual}), \qquad (1)$$

where NAV^{FV} is the value of the fund shares at the time the trading decision is made, which is usually assumed to be 4 PM, or a later time for a potentially late

trade. (1) is equivalent to the difference between the assets of the fund and what they would have been had the transactions been handled at fair value NAVs. The second version, used by Greene and Hodges (2002), captures the difference between the net cash injected into a fund by traders and what would have been injected had their trades been delayed until the next day:

$$\sum_t \Delta shares_t \cdot (NAV_{t+1}^{Actual} - NAV_t^{Actual}). \quad (2)$$

While the counterfactuals considered by the two versions are different, they usually yield similar results. If a fair value NAV is defined as the expectation of the next-day NAV given information available at 4 PM, then the difference between the two versions is the sum of net share purchases times the unexpected NAV change. So long as fund share trading is not conditioning on post-4 PM information, then this difference should be zero in expectation. The first (GIR-Z) version therefore yields a measure of dilution that varies less across time depending on whether post-4 PM market movements turned out to be advantageous for fund traders. On the other hand, the second (GH) version incorporates any late trading that conditioned on post-4 PM information, as well as any information that was available at 4 PM but omitted by the econometrician when constructing fair value NAVs.

The second, or "consulting", method was developed by Ciccotello and Greene (2006) and other expert witnesses as part of post-investigation consulting work. In addition to considering dilution from arbitrageurs trading at stale NAVs, the consulting method also considers the impact of arbitrageurs' flows on the portfolio of the fund. It is often implemented in practice by assuming that the proceeds from purchases from known short-term traders were held in cash and that this increased the cash holdings of the fund dollar-for-dollar. In periods of declining markets, such as the 2001 to early 2003 period in which much of the fund arbitrage covered by the cases occurred, increasing the cash content of a portfolio turned out to be helpful, at least *ex post*. The consulting method differs from the academic method in giving fund management companies credit for this presumed effect of arbitrage trading in increasing their cash holdings.

For example, suppose a short-term trader purchased a single fund share on day t and sold it on day $t+s$. The Greene and Hodges version of the academic method would calculate dilution from this transaction as

$$(NAV_{t+1}^{Actual} - NAV_t^{Actual}) - (NAV_{t+s+1}^{Actual} - NAV_{t+s}^{Actual}), \quad (3)$$

whereas the consulting method would calculate it as:

$$(NAV_{t+s}^{Actual} - NAV_t^{Actual}) - NAV_t^{Actual} \cdot [(1-\lambda) \cdot r^{Cash} \cdot s + \lambda \cdot (r_{t+1,t+s+1}^{Port})], \quad (4)$$

where the first term is the difference between the cash paid and received by the trader, and the second term is the return earned by the fund using the trader's money, assuming that a λ share is invested in assets normally held by the fund and

the remainder is held in cash. If $\lambda = 1$ and $(r^{Port}_{t+1,t+s+1}) = (NAV^{Actual}_{t+s+1} - NAV^{Actual}_{t+1}) \div NAV^{Actual}_{t}$, then the two methods will yield the same result. If $\lambda < 1$ and the mutual fund's portfolio underperformed cash ($r^{Cash} \cdot s > r^{Port}_{t+1,t+s+1}$), then the consulting method will yield a lower result. If $\lambda = 0$, then (4) equals the cash profits of the short-term trader. When dilution of shareholders and short-term trader profits are discussed as if they are identical, this is usually an indication that the consulting method is being employed with $\lambda = 0$.[13]

Whether one should credit a defendant for allowing arbitrage trading that ended up forcing it to hold more cash than it normally would have given its investment objective is debatable. Even if one accepts that one should, however, determining the effect of short-term trading on a portfolio's average cash position is far from straightforward. A reasonable theoretical model of the issue would be to assume that portfolio managers have an optimal percentage cash holding c^* and face convex losses $l(c-c^*)$ from deviations on either side. If this loss function is quadratic, the manager will target an average cash holding of c*. For example, faced with a short-term trader who invested $x in the fund on 50 percent of days, a manager who wanted to avoid incremental security trades would respond by targeting cash holdings of $c^* - x/2$ on days the short-term trader was out of the fund and $c^* + x/2$ on days she was in the fund, and the presence of the short term trader would not affect the average cash holdings of the fund.[14] So long as l' is not infinite below c^* or zero above c^*, managers should respond to a short-term trader by holding more cash when she is in the fund and *less* cash when she is out of the fund. The consulting method, as it is often implemented, ignores the latter effect, suggesting that it should overstate the effect of short-term trading on a fund's average cash holdings.[15]

---

[13] For example, in a frequently asked questions memo regarding its settlement with regulators, Waddell and Reed states, "as the Kansas Securities Commissioner's order recognizes, market timing can actually benefit a fund during a declining market. This is borne out by the fact that one of the market timers who paid service fees to Waddell & Reed actually lost more than $6 million while trying to time the funds. Nonetheless, the SEC ultimately determined that investors in aggregate profited from timing the fund and that such profits diluted the funds." Waddell and Reed, Memo, July 2006, 2 (http://media.corporate-ir.net/media_files/IROL/72/72585/news/072406.pdf, last accessed April 13, 2007).

[14] In this specific example, with a short-term trader who is in the fund 50 percent of the time, it is sufficient for the loss function to be symmetric for the manager to target c*.

[15] Even when the consulting method analyzes whether incremental short-term money was invested or held in cash, it assumes that funds did not target a lower cash holding when the short-term money was out of the fund. See, for example, Hamermesh (2006, p. 4): "For any given mutual fund, if the funds a timer uses to purchase fund shares were never invested in risky portfolio assets by the portfolio manager, timer net gains would accurately measure dilution to contemporaneous shareholders. To the extent, however, that timer funds were invested in risky

Indeed, an empirical analysis reveals that funds experiencing heavy short-term trading had average cash holdings that were *no higher* than other funds, consistent with what one would have expected from the discussion of loss functions above but inconsistent with the implicit assumptions of the consulting method. Regressions in Table 2 examine the correlation between short-term trading activity, as measured by fund share turnover, and both the average cash holdings and average betas of international funds.[16] It finds no evidence of the positive correlation between short-term trading and either cash holdings or betas that is often assumed by the consulting method, whether or not other variables that might influence average cash holdings are controlled for. The coefficients and standard errors imply that the upper limit of the 95 percent confidence interval equates to a 100 percent increase in fund share turnover being associated with a 12 basis point increase in the share of the portfolio held in cash or a beta that is 0.002 higher, so an economically meaningful positive correlation can be ruled out.

For comparison, the Table also presents identical regressions with returns as the dependent variable. In contrast to the lack of a relationship with cash holdings, fund share turnover is strongly negatively related to returns: 100 percent higher fund share turnover is associated with returns that are 13 basis points lower. This relationship is of essentially identical magnitude in years with rising and declining markets, casting further doubt on the idea that arbitrage trading helps funds in down markets by increasing cash holdings.

---

assets, dilution may diverge from timer net gains. Accordingly, it is necessary to estimate the extent to which portfolio managers invested in (and sold) portfolio assets in response to timer investment (and sales)."

[16] Cash holdings data are taken from CRSP, but results are virtually identical if one uses cash holdings data from Morningstar CDs.

Table 2. Fund share turnover, average cash holdings, and returns
General international equity funds, 2000-3

		Cash holdings in percent		Beta on EAFE index (x100)		Annualized returns in percent	
Independent variable	Defintion	(1)	(2)	(3)	(4)	(5)	(6)
Fund share turnover rate (t)	Min(sales,redemptions)/	0.047	0.032	0.035	0.050	-0.131**	-0.134**
	avg assets	(0.051)	(0.045)	(0.084)	(0.096)	(0.051)	(0.055)
Ln[Average monthly assets(t)]			-0.109		0.639**		-0.391
			(0.110)		(0.315)		(0.190)
Expense ratio (t) in percent	Percent		-0.713**		-0.530		-2.252***
			(0.305)		(1.047)		(0.653)
Portfolio turnover rate (t) in percent	Percent		0.348		1.527**		-2.696**
			(0.277)		(0.769)		(0.539)
Fixed effects		Year	Year	Year	Year	Year	Year
Observations		1,157	1,157	1,157	1,157	1,157	1,157

In this table, and throughout the paper, standard errors are heteroskedasticity robust and significance at the 10, 5, and 1 percent level is indicated by 1, 2, and 3 asterisks, respectively. The unit of observation is a fund*fiscal year combination. For the beta regressions, the dependent variable is a Scholes-Williams (1977) beta estimated for each fund calendar year on the EAFE index using daily data and one lead and lag.

Table 3. Interaction of dilution measurement methodology and time period
International equity funds

	Asset-weighted annualized dilution in percentage points, by method				
	Academic method		Consulting method		Average annualized
Year	GIR-Z	GH	Min cash target	Mean cash target	fund return
1998 (Feb to Dec)	0.29	0.33	0.67	0.49	13.68
1999	0.24	0.30	1.43	0.93	43.88
2000	0.40	0.34	0.30	0.44	-23.10
2001	0.45	0.39	0.17	0.39	-22.31
2002	0.47	0.46	0.23	0.45	-30.00
2003 (Jan to June)	0.33	0.36	0.23	0.20	28.77
Average (1998-2003)	0.36	0.35	0.54	0.50	1.82
1998-99	0.26	0.31	1.05	0.71	28.78
2001-03	0.42	0.42	0.23	0.37	-11.66
Quarterly standard deviation	0.17	0.14	1.12	0.56	13.10
Average in positive return quarters	0.30	0.29	1.21	0.77	36.90
Average in negative return quarters	0.43	0.43	-0.34	0.16	-46.47

IV. Constructing Fund-level Dilution Estimates from Publicly Available Data

Given this background, I shall proceed by calculating dilution using four different methods: the GH and GIR-Z versions of the academic method and two versions of the consulting method, one that assumes portfolio managers target a given average cash balance, and one that assumes managers target a minimum cash balance. The GH method is implemented as described in that paper. The GIR-Z method requires estimation of a fair value NAV, which is done assuming a transaction time of 9 PM as in Zitzewitz (2006) to take account of dilution due to any late trading. As in Zitzewitz (2003), Tufano (2005), and Zitzewitz (2006), a portfolio-level fair value NAV is estimated by regressing NAV returns on four predictive factors: the difference between the Chicago Mercantile Exchange Nikkei 225 future price at 4 PM Eastern time and the closing Nikkei futures price from the Singapore Exchange and S&P future returns from 3 to 11:30 AM, 11:30 AM to 4 PM, and 4 PM to 9 PM.[17] To avoid a data-snooping bias, the weights on these factors are estimated for each day using a rolling one-year window of prior data and then applied out of sample to the factor returns from that day. This methodology approximates that of commercial fair value pricing services that since 2003 have been increasingly used to value international equities.[18]

The consulting method requires estimating the fund share holdings of short-term traders and then making an assumption about how the portfolio manager reacts to these holdings when setting cash holdings. I approximate short-term trader holdings as the difference between the number of fund shares outstanding (adjusted for reinvestments of distributions) on day *t* and the minimum number of shares outstanding in the prior 20 trading days.[19] For manager responses to short-term traders, I make two relatively extreme

[17] The breakpoints for S&P futures changes are chosen to capture the approximate capitalization-weighted average closing time of Asian markets (3 AM), European markets (11:30 AM), U.S. equity markets (4 PM), and the latest time that late trades have been reported being placed (9 PM).

[18] See, for example, U.S. Patent No. 7167837, "Fair Value Pricing of a Financial Asset."

[19] For funds I have analyzed where I do have account-level data, this method has performed well in approximating the results of direct measurement of the holdings of short-term traders. One shortcoming of this method is that when a fund is growing, the method will incorrectly label some of the additional assets as short-term trader assets. On the other hand, it likely that a manager might not know the intentions of every trader that purchases his fund, and thus might face the same uncertainty. An alternative method, which assumes perfect foresight of shareholder holding periods by managers and measures short-term trader holdings as the difference between shares outstanding on day t and the higher of 1) the lowest shares outstanding from the prior 20 days and 2) the lowest shares outstanding from the next 20 days, yields average short-term trader holdings that are about half as large as with the no-foresight method. Given this, the no-foresight method probably overstates short-term trader holdings, and thus overstates any impact they would have on the cash holdings of a fund.

assumptions in the hope that they bound the truth. First, in the "min cash target" version, I assume that the cost to a manager of being below her target cash holdings are infinite, and thus managers target a minimum cash holding and hold all short-term trader funds in cash. Second, in the "mean cash target" version, I assume that managers target an average daily holding, so short-term trading leads does not lead to an increase in average cash holdings.[20] The first assumption is more consistent with how the consulting method is often implemented in practice, while the second assumption is more consistent with the lack of a correlation between fund share trading and cash holdings reported in Table 2.

Table 3 reports estimates of asset-weighted average annualized dilution rates for international equity funds by year, constructed using each of the four methods. Whereas most of this study will focus on the 2000-3 period emphasized by the regulators, in order to examine the effect of the general direction of the market on dilution estimates, in this table I include the up-market years 1998-9.[21] While estimates constructed using both versions of the academic method are fairly consistent regardless of the general direction of the market, the consulting method yields higher dilution estimates during the up market of 1998-99 and lower dilution estimates during the down market of 2001-3. As one would expect, the "min cash target" version of the consulting method, which assumes that short-term money increases the cash holdings of the fund, yields dilution estimates that are the most affected by the direction of the market. Indeed for small-cap equity funds (omitted in Table 3), the "min cash target" method actually yields *negative* dilution estimates for 2001-3. In contrast, the "mean cash target" version yields results that are more consistent with the academic methods.

All of these methods for measuring dilution require fund-level daily flow data, which are not publicly available. Past work on dilution uses proprietary daily fund flow data for a sample of funds and extrapolates from the sample to the industry as a whole (GIR, 2001; Greene and Hodges, 2002; Zitzewitz, 2003; Zitzewitz, 2006). The focus on industry-level dilution allows this work to avoid discussing observations for particular fund families.

---

[20] Specifically, I assume that the difference between the shares outstanding on day t and the difference between the average number of share outstanding of the last 20 days is held as extra cash. This implies that a fund will hold more cash when short-term traders have recently bought the fund and less cash when they have recently sold.

[21] My two sources of data asset data are Lipper (March 2000 to December 2003) and TrimTabs (February 1998 to December 2003). For consistency between years in Table 3, I use only the TrimTabs data. Adding the Lipper data after March 2000 does not significantly affect results, however.

This paper, in contrast, requires estimating family-level dilution for comparison with settlement amounts, including fund families for which daily flow data are not available, and so I must take a different approach. I exploit the relatively close correlation between publicly available data on fund share turnover and dilution. For each category of fund and each year (e.g., Asia/Japan) and each year, I estimate the relationship between fund share turnover and dilution, using the data for the sample of funds for which I have daily flow data from data flow data provided to me by Lipper and Trim Tabs (this joint Lipper and Trim Tabs sample is also used in Zitzewitz, 2006). I then use these relationships to generate predicted dilution rates for every fund, using the fund's publicly reported fund share turnover. This approach therefore utilizes individual funds' daily flow data only to estimate the overall dilution-turnover relationship, and these data are not used in the dilution estimates for any given fund family.

The procedure starts by matching gross fund share sales and redemption by fund and fiscal year for 2000 to 2003 to either average daily asset data from Lipper or average month-end asset data from CRSP (in order of priority) using ticker symbols. Fund share turnover can thus be calculated for every fund with a ticker symbol for which Lipper collected gross sales and redemption data from the original SEC filings. The second step is to use the combined Lipper and Trim Tabs daily flow data (described in more detail in Zitzewitz, 2006) to estimate dilution rates for fund*fiscal year combinations and to match these dilution rates with the fund turnover rates. Predicted dilution rates are then estimated from fund share turnover rates using the semi-parametric method described below, and predicted dilution is then summed across a company's funds to yield a company-level measure. The process is repeated for each of the four dilution measurement methods discussed above (i.e., the GIR-Z and GH versions of the academic method, and the consulting method assuming that managers target minimum or average cash holdings).

Table 4 reports summary statistics for the sample of international equity funds in CRSP, for the subset of these which can be matched with Lipper fund share turnover data, and for the smaller subset for which daily flow data is also available. An observation is a fund*fiscal year. Given that most funds have fiscal years that end in months other than December (October is the most common), the annual variables taken from CRSP (expense ratios, loads, portfolio turnover, cash holdings) are from the prior calendar year. Average month-end assets are calculated using the data for the exact months of the funds' fiscal year. Fund share turnover data is available for 57 percent of fund*years and 80 percent of assets; about 62 percent of these observations (and 38 percent of these assets) also have daily flow data that allows the calculation of dilution rates. Fund share turnover is calculated as the minimum of sales and redemptions divided by average month

end assets.[22] As others have noted (e.g. Greene and Hodges, 2002), the ultimate source for the sales and redemption data, Form N-SAR filings, often include figures that are off by 3 orders of magnitude due to filers entering figures as whole dollars or millions instead of thousands. This results in some obvious outliers; to limit their impact, I treat observations with fund share turnovers below 0.1 or above 50 as missing.[23]

A comparison of the difference between the equal and asset-weighted averages reveals that fund share turnover and dilution are negatively correlated with fund size. Apart from differences in average fund size, the sample of funds with share turnover information is roughly representative of the broader sample, as is the subset of those funds with daily flow data. The equal-weighted average fund share turnover rates and dilution rates are about 2.5 and 1.5 percent for this latter subsample, respectively. The ratio of these two is informative about the extent to which trading in the fund was opportunistically timed. A fund share turnover ratio of 2.5 implies that a fund with a $100 in average daily assets would have experienced at least $250 each in fund share sales and redemptions in over the course of a year (compared with $66 for non-international funds). Dilution of 1.5 percent implies that these $250 in roundtrip trades were accompanied by $1.50 in dilution, or about 0.6 percent of the roundtrip trading volume. This could have resulted, for example, from these roundtrip trades involving purchases when the fund was underpriced by 0.3 percent and sales when it was overpriced by an equivalent amount. In comparison, the average daily absolute difference between the stale and fair value NAVs for these funds was 0.35 percent. This suggests that most trades in and out of these funds were opportunistically timed.[24]

[22] An alternative would be to divide the *average* of fund share sales and redemptions by average assets, which would yield a higher measure for a fund that was growing or shrinking over the course of the year. I use the minimum of sales and redemptions as it yields a slight better predictor of fund share trading that is correlated with NAV predictability and thus produces dilution.

[23] These cutoffs are arbitrary, but results are not very sensitive to them. It should be noted that when average daily assets are used as the denominator, the maximum possible fund share turnover in a typical year with 252 trading days is 126, assuming T+1 settlement of fund shares.

[24] In instances where I have had the opportunity to analyze account-level data for funds affected by arbitrage, it has always been the case that arbitrageurs, who trade much more actively than typical shareholders, accounted for a small percentage of a fund's assets but a large percentage of a fund's trading activity.

Table 4. Summary statistics

	All international equity funds in CRSP with tickers	All fund years with share turnover data		All fund years with share turnover and daily flow data	
	Equal weighted	Equal weighted	Asset weighted	Equal weighted	Asset weighted
Observations	5,525	3,153		1,965	
Unique funds	1,741	1,137		819	
Average month-end assets ($ millions)	192	268		165	
Expense ratio (percent)	1.89	1.86	1.17	1.90	1.32
Load fund dummy	0.66	0.66	0.53	0.68	0.42
Portfolio turnover	1.03	1.00	0.61	1.01	0.73
Cash holdings (per CRSP)	3.8	3.8	5.0	3.8	4.5
Fund share turnover		2.01	0.70	2.08	0.99
Dilution rates by method (Percent)					
Academic (GIR-Z)				1.58	0.62
Academic (GH)				1.33	0.59
Consulting (Min cash target)				0.92	0.45
Consulting (Mean cash target)				1.18	0.50

Data from CRSP include international equity funds (ICDI objective = international equity) for the calendar years 1999-2002. These are matched using tickers with Lipper data for the following fiscal year on fund share sales and redemptions and on dilution rates calculated from daily flow rates.

Table 5. Asset-weighted average dilution rates by asset class and fund share turnover

	Annualized dilution rate (asset-weighted average, GIR-Z method), by asset class							
	International equity funds				Global/Latin			
Fund share turnover rate	General intl. equity	Emerging Market	Asia/Japan	Europe	American equity funds	Small/midcap equity funds	Total	No. of obs.
0.1 to 0.19	0.01	0.00	0.12	0.06	0.07	0.01	0.03	1,524
0.2 to 0.49	0.08	0.04	0.16	0.10	0.08	0.01	0.04	2,654
0.5 to 0.99	0.40	0.38	0.88	0.68	0.20	0.09	0.25	1,002
1 to 1.9	0.69	0.83	1.41	0.92	0.42	0.40	0.60	584
2 to 4.9	2.44	1.74	4.30	1.76	1.34	1.30	1.84	384
5 to 9.9	5.34	4.69	5.02	11.30	1.70	2.19	3.42	118
10 to 49	18.17	6.63	8.31	12.37	6.91	4.69	15.36	111
All observations	0.59	0.38	1.96	0.92	0.13	0.08	0.22	6,377
Number of observations	1,319	246	235	165	534	3,878	6,377	

Notes: An observation is a fund*fiscal year combination. Data is for fiscal years 2000-3. Fund share turnover is defined as the minimum of the value of fund share sales and fund share redemptions, divided by average month-end assets

Table 5 examines the relationship between fund share turnover and dilution in more detail. Fund*years are divided into cells based on their amount of fund share turnover and their rough asset class.[25] Average dilution is clearly increasing in fund share turnover. In addition, the regionally focused European and Asian stock funds experience more dilution and more dilution for a given fund share turnover rate than general international and emerging market equity funds, which in turn experience more dilution than global or domestic small and midcap equity.[26] Given this, when econometrically estimating the relationship between share turnover and dilution, I will separate domestic small/midcap equity, global equity, diversified (international and emerging market) and regionally focused (Europe and Asia) funds.

Table 6 presents regressions that use fund share turnover to predict dilution. In Panel A, average turnover rates, dilution rates, and their ratio are reported for diversified and regionally focused funds for different levels of share turnover. In panel A, it appears that funds with low fund share turnover rates also have, understandably, very low dilution. Dilution rises more or less linearly with fund share turnover, although the dilution-turnover ratio does appear to be maximized at intermediate levels of dilution. Given the approximate linearity of the relationship, I first estimate an asset-weighted linear predictive model. This appears to estimate negative dilution for funds with very low turnover rates, which is undesirable, so I instead estimate linear spline models, with the kinks at the 20th, 40th, 60th, and 80th percentile of the overall fund share turnover rate distribution. These models offer a modest improvement in R-squared.

In the analysis that follows in the rest of the paper, I estimate the spline regressions separately for each year, each category, and for each of the four dilution methodologies, and I construct a predicted dilution rate for each fund*year for which I have fund share turnover data. For the fund*years where I lack this data, I assume a dilution rate equal to the asset-weighted average for that year and asset class. I then add dilution across a mutual fund companies' funds to construct an estimate of fund-family dilution for the fiscal years 2000-3 period. It is this sum that I compare with reimbursements and penalties to evaluate the outcome of settlement negotiations.

---

[25] I classify funds by asset class using information, in order of priority, from Lipper, Morningstar (matched by ticker), and CRSP (matched by ticker). In Table 5, the categories for Asian Stock funds, Japan Stock funds, Chinese stock funds, and Pacific (x. Japan) stock funds are aggregated, as are the categories for global and Latin American equity (which have similar dilution levels).

[26] As Zitzewitz (2003) discusses, these funds attracted more arbitrage trading because trading one of either Europe or Asia offered more arbitrage profits than trading a fund that was a combination of the two.

Table 6. Fund share turnover (FST) and dilution, by asset class

Panel A. Average fund share turnover and dilution rates in percent (asset-weighted)

	Gen. Intl. Equity and Emerg. Mkt. funds			Europe, Japan, and Asia equity funds			Global and Latin America funds			Small and midcap domestic equity funds		
FST category	FST	Dilution	Dil/FST	FST	Dilution	Dil/FST	FST	Dilution	Dil/FST	FST	Dilution	Dil/FST
0.1 to 0.19	0.15	0.04	0.29	0.16	-0.01	-0.05	0.15	0.00	0.03	0.15	0.00	-0.01
0.2 to 0.49	0.32	0.12	0.36	0.33	0.09	0.26	0.32	0.16	0.49	0.32	0.02	0.06
0.5 to 0.99	0.72	0.42	0.58	0.74	0.60	0.81	0.70	0.26	0.37	0.68	0.08	0.12
1 to 1.9	1.42	0.92	0.65	1.43	1.34	0.93	1.33	0.44	0.33	1.33	0.21	0.16
2 to 4.9	3.09	2.13	0.69	3.09	2.84	0.92	3.21	1.36	0.42	2.89	0.51	0.18
5 to 9.9	6.84	5.85	0.86	6.50	7.17	1.10	6.72	3.27	0.49	7.08	1.96	0.28
10 to 49	19.86	13.60	0.68	17.70	15.98	0.90	19.09	4.39	0.23	17.51	3.26	0.19

Panel B. Regressions predicting dilution from FST

	Linear model	Linear spline model	Linear model	Linear spline model	Linear model	Linear spline model	Linear model	Linear spline model
Constant	-0.33***	-0.05	-0.02	0.10	0.01	0.02	-0.09***	-0.01
	(0.10)	(0.05)	(0.11)	(0.12)	(0.03)	(0.06)	(0.01)	(0.02)
FST	0.95***	0.50	0.82***	-0.25	0.32***	0.28	0.37***	0.09
	(0.11)	(0.32)	(0.07)	(0.99)	(0.07)	(0.35)	(0.05)	(0.14)
FST*(FST > 20th pctile)		-0.23		-0.92		0.23		0.13
		(0.38)		(1.04)		(0.42)		(0.10)
FST*(FST > 40th pctile)		1.65***		2.91**		-0.08		-0.02
		(0.52)		(1.20)		(0.34)		(0.13)
FST*(FST > 60th pctile)		0.08		0.72		0.64***		0.44**
		(0.29)		(0.52)		(0.19)		(0.19)
FST*(FST > 80th pctile)		1.01***		0.80***		0.28***		0.39***
		(0.12)		(0.08)		(0.09)		(0.04)
Observations	1565		332		534		3878	
R squared	0.8292	0.8393	0.6761	0.6783	0.3887	0.3981	0.6636	0.6727

Notes: The 20th, 40th, 60th, and 80th percentile cutoffs are 0.23, 0.45, 0.98 and 2.60 across all asset classes. An observation is a fund*fiscal year combination. Observations are asset-weighted. Standard errors adjust for clustering within fund.

The R-squareds in these spline regressions indicate that fund share turnover explains most of the variation in dilution rates across funds (83 and 68 percent, respectively, for general international equity and Europe/Japan/Asia funds, which together account for at least 80 percent of dilution). This is important to note because the approach in this paper of estimating dilution from fund share turnover assumes that the dilution-turnover relationship is the same across funds in the same year and category. The fact that the R-squareds are not 100 percent indicates that this is not exactly true: in some funds trades were more opportunistically timed, and thus caused more dilution per dollar transacted. Fortunately, the relationship between predicted and actual dilution becomes even tighter once we aggregate to the family level. For the fund families for which I can calculate actual dilution, predicted the dilution-to-assets ratio in a family's international equity funds explains 92 percent of the variation in the actual dilution-to-assets ratio.

V. Analysis of the Settlements

Table 7 provides data on the 20 settlements of market timing and late trading cases negotiated with mutual fund families as of January 2007.[27] Most settlements were negotiated in 2004. The settlements usually involved simultaneously announced settlements of federal and state regulatory investigations. Sixteen of the 20 settlements involved the state of New York, and seven of those involved another state as well. All settlements included a restitution component, and 18 of the 20 included a penalty as well. Once restitution and penalties are agreed upon, they are distributed to affected shareholders according to a plan developed by an Independent Distribution Consultant (IDC), an expert who is selected by the board of the affected funds and whose choice must be approved by the regulators.

[27] The regulators have also brought cases against brokers and hedge funds, for example, for falsifying records and engaging in late trading. Since the dilution from these cases occurred in multiple families' funds, I cannot construct a dilution estimate for each settlement and therefore omit them from my analysis.

Table 7. Settlements of market timing and late trading cases negotiated with fund companies,2003-7

			Settlement amount ($ millions)			Case characteristics	
Family	Settlement Date	States involved	Restitution	Penalty	Fee Reduction	Late trading involved	Senior manager traded in own funds
Alliance	December 18, 2003	NY	150	100	350	0	0
MFS	February 05, 2004	NY, NH	175	50	125	1	0
Bank of America	March 15, 2004	NY	250	125	160	1	0
Fleet	March 15, 2004	NY	70	70	0	0	0
Putnam	April 08, 2004	MA	10	100	0	0	1
Janus	April 27, 2004	NY, CO	50	50	125	0	0
Strong	May 20, 2004	NY, WI	70	70	35	0	1
Pimco (PEA)	June 01, 2004	NJ	1.6	18	0	0	0
Pilgrim Baxter	June 21, 2004	NY	160	110	10	0	1
Banc One	June 29, 2004	NY	10	40	40	0	0
Franklin	August 02, 2004	MA	30	20	0	0	0
RS	October 06, 2004	NY	12	14	0	0	0
AIM	October 08, 2004	NY, CO	20	30	0	0	0
Invesco	October 08, 2004	NY, CO	210	115	75	0	0
Fremont	November 04, 2004	NY	2	2	0	1	0
Federated	November 28, 2005	NY	35	45	20	0	0
AXP	December 01, 2005	MN	15	0	0	0	0
Waddell & Reed	July 24, 2006	NY, KS	50	2	25	0	0
Scudder	December 22, 2006	NY, IL	102	20	86	1	0
Alger	January 18, 2007	NY	30	10	5	0	0
Totals		20	1,453	991	1,056		
NY involved		16	1,396	853	1,056		
NY not involved		4	57	138	0		

Source: SEC and state press releases. For settlements involving additional issues such as fund selling practices (e.g., American Express), figures include only those identified as being for market timing/late trading. Settlements or fines levied by self-regulatory bodies such as the NASD and settlements with individuals are excluded, except where the individual is a firm owners (e.g., Strong).

In addition to restitution and penalties, the New York Attorney General also negotiated future fee reductions as part of 12 of the settlements. The SEC has taken the position that it should not regulate mutual fund fees and thus, while most aspects of the settlements have been coordinated between the SEC and the states, the SEC has explicitly noted that it did not participate in the fee negotiation part of the settlements. Determining the extent to which a fund company that is forced to reduce its fees has been punished is, of course, quite complicated.[28] For simplicity, I will simply ignore the fee reduction component of the settlement and simply analyze restitution and penalties. Since only New York insisted on fee reductions, excluding them will tend to make New York-negotiated settlements appear smaller.

Table 8 compares the settlement amounts to my estimates of family-level dilution for the fiscal years ended 2000-3, as well as to other measures of economic size and trading activity. The restitution-to-estimated dilution ratio ranges from 0.04 to 5; the four settlements not involving New York include the first, second, and third-lowest ratios. Restitution-to-dilution ratios are very similar to ratios with alternative denominators, either total mutual fund assets, international equity fund assets, or the latter multiplied by average fund share turnover rate, foreshadowing the fact that conclusions about an "Eliot effect" are not very sensitive to the method used to measuring dilution. The average estimated dilution rate (using the GIR-Z version of the academic method) is reported; it is understandably higher for families with a higher average fund share turnover rate and for families with a higher proportion of regionally-focused funds.

[28] One issue is that if fees begin at a profit-maximizing level, then, at that level, the derivative of profits with respect to fees should be zero and thus the first increment of fee reduction should not be costly at all. In addition, if fund firms are tacitly colluding on fees, a regulator-mandated fee reduction may benefit a fund firm by allowing it means of reducing price without generating a competitive response.

Table 8. Settlement restitution compared with estimated dilution and family size

			Estimated dilution, 2000-3 ($ millions)		Assets, end of 2003 ($ billions)		Asset-weighted averages, international equity funds		Ratios (Restitution divided by:)				
Family	States involved	Restitution ($ millions)	Intl, Global, Small, Midcap	Intl only	Total	Intl. Equity	Fund share turnover	Predicted dilution rate (percent)	Dilution	Intl. Eq. Dilution	Intl. Eq. Assets* turnover (percent)	Intl. Eq. Assets (percent)	Total Assets (percent)
MFS	NY, NH	175	35	16	80	2.6	0.91	0.49	5.07	11.27	6.12	6.69	0.22
Strong	NY, WI	70	14	8	28	0.3	3.43	1.85	4.99	8.38	95.71	27.92	0.25
Alliance	NY	150	64	37	112	8.3	0.33	0.15	2.33	4.03	0.59	1.82	0.13
Bank of America	NY	250	117	94	132	6.0	1.06	0.70	2.14	2.66	4.39	4.15	0.19
Invesco	NY, CO	210	150	22	104	0.4	2.25	0.94	1.40	9.67	133.41	59.29	0.20
Waddell	NY, KS	50	45	37	20	0.8	1.49	0.89	1.12	1.34	9.37	6.27	0.25
Fleet	NY	70	89	73	96	4.7	1.00	0.66	0.79	0.96	1.49	1.49	0.07
AXP	MN	15	20	12	78	1.7	0.65	0.32	0.74	1.28	0.56	0.87	0.02
Alger	NY	30	43		4	0.0			0.69				0.68
Pilgrim Baxter	NY	160	282	67	21	4.3	0.92	0.47	0.57	2.39	3.38	3.68	0.77
Banc One	NY	10	20	14	101	1.8	0.43	0.23	0.49	0.71	0.24	0.56	0.01
Scudder	NY, IL	102	268	226	126	5.5	1.32	0.78	0.38	0.45	2.45	1.86	0.08
Federated	NY	35	96	77	166	1.4	1.77	1.07	0.36	0.46	4.52	2.56	0.02
Janus	NY, CO	50	255	141	114	4.9	0.91	0.49	0.20	0.35	0.93	1.02	0.04
RS	NY	12	70		10	0.0			0.17				0.12
Fremont	NY	2	19	9	3	0.05	6.43	5.04	0.11	0.21	28.56	4.44	0.06
AIM	NY, CO	20	228	197	34	2.7	1.88	1.34	0.09	0.10	1.39	0.74	0.06
Franklin	MA	30	479	279	194	24.7	0.74	0.39	0.06	0.11	0.09	0.12	0.02
Putnam	MA	10	246	205	147	12.6	0.70	0.34	0.04	0.05	0.06	0.08	0.01
Pimco (PEA)	NJ	1.6	41	18	141	0.2	3.51	3.03	0.04	0.09	2.67	0.76	0.00
Total settlements		1,453	2,583	1,532	1,712	82.8	0.87	0.49	0.56	0.95	1.53	1.75	0.08
NY involved		1,396	1,795	1,018	1,151	44	0.99	0.59	0.78	1.37	3.17	3.20	0.12
NY not involved		57	787	514	560	39.2	0.74	0.39	0.07	0.11	0.11	0.14	0.01

Table 9. Regressions explaining variation in restitution ratios across settlements
Dependent variable: Ln(Restitution or Restitution+Penalty) - Ln(Dilution)

#	Dependent variable	Observations (F-test p-value)	Constant	Late trading	manager trading in own funds	NY involved	NY only state involved	NY + other states involved	LN(Total mutual fund assets, end 2003)
1	Restitution	20	-1.09***	0.91	0.28				
		0.517	(0.39)	(0.78)	(1.33)				
2	Restitution + Penalty	20	-0.30	0.49	0.69				
		0.647	(0.32)	(0.71)	(0.92)				
3	Restitution	20	-2.39***			1.97**			
		0.015	(0.66)			(0.73)			
4	Restitution + Penalty	20	-1.04**			1.21**			
		0.021	(0.37)			(0.48)			
5	Restitution	20	-2.50***	0.44	0.43	1.89**			
		0.157	(0.79)	(0.80)	(0.93)	(0.84)			
6	Restitution + Penalty	20	-1.24**	0.17	0.80	1.26**			
		0.171	(0.49)	(0.74)	(0.76)	(0.56)			
7	Restitution	20	-2.50***	0.46	0.42		1.68*	2.13**	
		0.260	(0.81)	(0.80)	(0.88)		(0.87)	(1.01)	
8	Restitution + Penalty	20	-1.24**	0.18	0.79		1.18*	1.35*	
		0.290	(0.50)	(0.76)	(0.74)		(0.62)	(0.71)	
9	Restitution	20	-4.80	0.55	0.44	2.09**			0.12
		0.269	(3.21)	(0.75)	(1.00)	(0.97)			(0.16)
10	Restitution + Penalty	20	-3.57	0.28	0.81	1.46**			0.12
		0.300	(2.66)	(0.71)	(0.82)	(0.68)			(0.14)

Notes: Each row is a regression of the log restitution ratio on explanatory variables. Standard errors are heteroskedasticity robust; significance at the 10, 5, and 1 percent level is indicated with 1, 2, and 3 asterisks, respectively.

Table 9 presents regressions that seek to econometrically explain the variation in restitution ratios. The log restitution ratio (with and without penalties) is regressed on variables capturing the severity of conduct, the involvement of New York state, and the fund firm's ability to pay. The regressions find higher ratios for cases that involved late trading or where senior managers were alleged to have conducted arbitrage trading in their own funds, but neither difference is statistically significant. Ratios are (statistically and economically) significantly higher in cases involving New York by between 1.2 and 2.1 log orders of magnitude (i.e. by a factor of between 3.3 and 8.4). Regressions that test whether settlements involving only New York differed from those involving New York and other states find no significant differences. Adding a proxy for ability to pay, the fund family's log mutual fund assets at the end of 2003, does not affect results. The constant in all but the last two regressions has a natural economic interpretation as the log restitution ratio for cases that do not involve late trading, senior manager trading, and/or New York. When the null hypothesis of a zero constant is rejected, this is equivalent to rejecting the hypothesis that the restitution ratio in these cases was one.

Table 10 presents instrumental variables (IV) regressions aimed at testing whether the correlation between restitution ratios and New York's involvement reflects a causal effect. In this context, a valid instrument would be a variable that affects the likelihood that New York would become involved in a case, but is uncorrelated with any unobserved factor affecting the restitution ratio, such as the severity of conduct. State regulators generally have a rationale for involving themselves in a securities case to the extent that the conduct in question took place in their state or to the extent that residents of their state were harmed by it. This suggests two possible instruments for New York's involvement: whether a fund firm is headquartered in New York and the share of its investors who are New York residents.

Table 10. Instrumental variables regressions
Dependent variable: New York Involvement, Ln(Restitution/Dilution) or Ln((Restitution+Penalty)/Dilution)

Panel A. First stage and reduced form regressions				Instrumental variables		
Dependent variable	Specification	Observations (R-squared)	Constant	NY share of single-state muni bond fund assets	No single-state muni fund dummy	NY headquarters?
New York involved	First stage (linear probability)	20	-2.09*	3.07**	-0.12	0.54**
		0.3064	(1.09)	(1.19)	(0.21)	(0.27)
Restitution	Reduced form	20	-14.32***	15.51***	-2.35***	1.84**
		0.3102	(3.84)	(4.29)	(0.77)	(0.84)
Restitution + Penalty		20	-7.85**	8.96**	-1.36**	0.88
		0.1489	(3.34)	(3.73)	(0.69)	(0.73)
Panel B. Instrumental variables						
Dependent variable	Specification		Constant	NY involvement		
Restitution	OLS	20	-2.39***	1.97***		
		0.284	(0.66)	(0.73)		
Restitution+Penalty		20	-1.04***	1.21**		
		0.1651	(0.37)	(0.48)		
Restitution	IV	20	-2.73***	2.39***		
		0.2713	(0.63)	(0.70)		
Restitution+Penalty		20	-1.09**	1.27**		
		0.1647	(0.46)	(0.56)		
Restitution	IV (using muni bond IVs only)	20	-1.66	1.05		
		0.2222	(2.09)	(1.69)		
Restitution+Penalty		20	-0.69	0.78		
		0.1442	(1.25)	(1.60)		

Notes: Hausman tests do not reject the equivalence of OLS and IV or IV and IV (muni bond IV only). No p-value in these tests is smaller than 0.2. Of the 20 settlement firms, 3 have New York Headquarters (Alger, Alliance, and Scudder) and 9 offer single-state muni bond funds (Alliance, American Express, Bank of America, Federated, Fleet, Franklin, MFS, Putnam, and Scudder). For these nine, the average NY asset share is 0.15 with a standard deviation of 0.08. Standard errors are heteroskedasticity robust; significance at the 10, 5, and 1 percent level is indicated with 1, 2, and 3 asterisks, respectively.

The anecdotal evidence I have been able to obtain about why New York was not involved in four settlements (American Express, Franklin, Pimco, and Putnam) is consistent with headquarters location being important. The Putnam case arose from a Massachusetts-resident internal whistleblower who contacted the Massachusetts regulator. In the Pimco and American Express cases, the home-state regulator (New Jersey and Minnesota, respectively) played a lead role. Franklin, which is headquartered in California but whose settlement involved Massachusetts, is the exception where headquarters does not appear to have played a role. I was told by an individual familiar with the matter that New York did not involve itself in these four cases because the other states expressed an interest in handling them without New York's involvement, perhaps to avoid the perception that they were playing a secondary role.[29]

The identifying assumption involved in using headquarters location and investors' residency as instruments for New York's involvement would be that mutual fund firms that are headquartered in New York, or have more New York resident customers, did not have higher settlement amounts for unobserved reasons, such as having engaged in more egregious conduct or having less bargaining power when negotiating with regulators. Two of the hedge funds first identified as having engaged in arbitrage trading were from the New York area (Canary Capital of New Jersey and Millennium Partners of New York), raising the possibility that conduct in New York-based mutual fund firms was more egregious. Perhaps surprisingly though, on the whole New York-based hedge and mutual fund firms were actually slightly less likely to be involved in these investigations than average. Four of the 20 mutual fund defendants are based in the New York area, compared with about 30 percent of all U.S.-based mutual funds (see, e.g., Massa, Reuter, and Zitzewitz, 2009). Likewise, three out of eleven U.S.-based hedge fund defendants from SEC litigation releases involving market timing and late trading are based in the New York area, compared with 60-70 percent of all U.S.-based hedge funds. In addition, of the seven settlements involving either late trading or senior manager trading, only one (Scudder) involved a New York-area firm.[30]

---

[29] This claim suggests that the non-New York cases were the result of other states having selected the cases (and requested a lead role), rather than New York having passed on them. Assuming that all states find cases more attractive if higher restitution ratios are likely, the sign of any selection bias to the New York involvement coefficient should depend on whether New York or the other states do the selecting. If New York's non-involvement in these four cases was indeed the result of decisions by the other states, this might mitigate any concern that New York's involvement was correlated with unobservable characteristics that contributed to high restitution ratios.

[30] It has been suggested to me by a referee that proximity to Eliot Spitzer may help explain apparently better conduct of New York area firms. Alternatively, New York area firms may have

Table 10 presents first-stage, reduced form, OLS, and IV regressions. The proportion of mutual fund investors who are New York residents is not publicly available, so I use as a proxy the share of a fund family's single-state municipal bond fund assets accounted for by New York.[31] The first-stage regression reveals that settlements with fund families with a higher share of New York investors and with headquarters in New York were more likely to involve the New York Attorney General.[32] A comparison of the OLS and IV coefficients on New York involvement reveals that the latter are modestly larger, although Hausman tests do not reject the hypothesis that they are the same. Using only the municipal bond instrumental variables yields results that lose statistical (but not economic) significance; a test of the overidentifying restriction does not reject, suggesting that headquarters location is not endogenous. Using the New York headquarters instrument alone unfortunately does not yield precisely estimated coefficients.

Instrumental variables suffers from known small-sample biases (Nagar, 1959). Supplementary evidence is available from the reduced form regressions, which reveal that firms headquartered in New York and with a high share of New York single-state muni bond shareholders paid significantly higher restitution ratios (differences in ratios that include penalties are about half as large and are not statistically significant). Taken together, the regressions in Table 10 imply that firms with greater exposure to New York state regulation (via their headquarters or location or customer base) were more likely to have New York involved in negotiations, and paid higher restitution ratios.

Table 11 presents some alternative dilution measures and specifications as robustness checks. For international and global equity funds, switching dilution measurement methods does not affect ones conclusions about the effect of New York's involvement in settlements.[33] One might have expected regulators to

---

had access to higher quality legal and compliance advice. If the conduct of New York area defendants was less egregious in ways uncaptured by observables, since these defendants were more likely to be prosecuted by New York rather than other states, this would bias my results against finding a large New York effect.

[31] Single-state municipal bond funds hold municipal bonds issued by a particular state. They generally offer lower yields than other municipal bond funds, but are attractive to residents of that state because their interest is deductible from that state's income tax. For fund families that offer no single-state muni bond funds, I code the New York percentage variable as zero and the "no single state muni funds" variable as one. Note that the state resident mix of a fund family's muni fund investors need not be the same as the mix of investors affected by arbitrage, which occurred primarily in other asset classes. For this proxy to be useful as an instrument the two New York shares need to be positively correlated and the difference between them needs to be unrelated to unobserved factors affecting the restitution ratio.

[32] I follow Angrist and Krueger's (2001) advice and use a linear probability model for the first stage.

[33] The downward bias of the consulting method in down markets is large enough that including small cap equity funds leads the "min cash target" version of the method to produce negative

negotiate settlements in cases with the worst conduct first, to obtain higher restitution ratios in these cases, and for New York to have been more likely to have been involved in these case. Adding controls for settlement timing and a dummy for whether the settlement was announced after Christopher Cox replaced William Donaldson as SEC chairman does little to affect the coefficient on New York's involvement, however.[34] Likewise, dropping the Alger settlement, which was announced two weeks after Andrew Cuomo replaced Eliot Spitzer as New York attorney general, does not meaningfully affect the results.

Table 11 also considers the robustness of the results to alternative interpretations of circumstances surrounding the settlements. Two of the settlements, those with Bank of America and Fleet, were negotiated while the two firms were conducting merger talks. The time pressure created by an impending merger might have increased regulators' bargaining power. In addition, Bank of America was also accused of facilitating late trading as a broker in addition to as a fund company, and so the settlement might have been larger because it covered that alleged behavior as well. Although these two settlements had among the higher restitutions ratios and involved New York, excluding them does not materially affect the New York involvement coefficient.

---

dilution estimates for 17 of the 20 settlements. Since a restitution ratio with a negative denominator is difficult to interpret, I exclude small cap equity funds from the analysis of robustness to method.

[34] The fact that the transition from Chairman Donaldson to Cox seems not to have affected restitution ratios may be mildly surprising given public perceptions of the differences in the posture the two took towards the mutual fund industry. A possible explanation is that New York was involved in most of the cases. If New York was more aggressive than the SEC under either Chairman, and if settlement sizes were determined by the more aggressive regulator, this would explain the lack of an effect of the identity of the SEC Chairman.

Table 11. Robustness of New York coefficient
Dependent variable: Ln(Restitution or Restitution+Penalty) - Ln(Dilution)

		OLS		IV	
Panel A. Robustness to alternative denomiators					
Denominator	Asset classes included	Restitution	Restitution+Penalty	Restitution	Restitution+Penalty
Baseline	IE, GE, SC	1.97***	1.21**	2.39***	1.27**
[Dilution (Academic method, GIR-Z method)]		(0.73)	(0.48)	(0.70)	(0.56)
Dilution (Academic method, GIR-Z)	IE, GE	2.16***	1.40**	3.01***	1.97**
		(0.77)	(0.59)	(0.87)	(0.79)
Dilution (Academic method, GH)	IE, GE	2.21***	1.46***	3.00***	1.96**
		(0.76)	(0.55)	(0.89)	(0.77)
Dilution (Consulting method, Min cash target)	IE, GE	1.87**	1.13*	2.71***	1.69*
		(0.75)	(0.61)	(0.82)	(0.92)
Dilution (Consulting method, Mean cash target)	IE, GE	2.11***	1.35**	2.84***	1.80*
		(0.77)	(0.61)	(0.87)	(0.90)
Assets (end 2003)	All	3.63***	2.87***	6.10**	4.98*
		(0.66)	(0.48)	(2.32)	(2.33)
Assets (end 2003)	IE	2.60***	1.85**	3.70***	2.66***
		(0.69)	(0.70)	(0.85)	(0.75)
Assets*Fund Share Turnover	IE	2.39**	1.63**	3.04***	2.00**
		(0.70)	(0.49)	(0.84)	(0.74)
None		1.69**	0.93*	0.85	-0.27
		(0.66)	(0.51)	(0.66)	(0.66)
Panel B. Robustness to specification and sample variations					
Specification/sample variation					
Settlement date, Chairman Cox dummy	IE, GE, SC	2.33***	1.55***	4.04***	3.05***
		(0.56)	(0.41)	(1.29)	(1.16)
Excluding Alger	IE, GE, SC	1.97***	1.23**	2.45***	1.45**
		(0.75)	(0.50)	(0.78)	(0.66)
Excluding Fleet and Bank of America	IE, GE, SC	1.87**	1.12**	1.64	0.83
		(0.75)	(0.50)	(0.96)	(0.67)
Alternative restitution figure for Putnam	IE, GE, SC	1.41**	1.07***	1.03	0.93
		(0.70)	(0.50)	(0.67)	(0.66)
Dilution-weighting of observations	IE, GE, SC	2.03***	1.31***	2.14***	1.06**
		(0.38)	(0.50)	(0.45)	(0.53)

Note: Each cell is the regression coefficient on New York's involvement from OLS and IV regressions as in Table 10. Asset class codes are: IE = international equity, GE = global equity, SC = small and midcap equity. Standard errors are heteroskedasticity robust; significance at the 10, 5, and 1 percent level is indicated with 1, 2, and 3 asterisks, respectively.

A second issue is which amount to use for the settlement with Putnam. The Putnam settlement was unique in that the restitution amount was a minimum that could be subsequently adjusted upward if the IDC found that actual harm to shareholders had been higher.[35] As it turned out, in his report as Putnam IDC, Peter Tufano (2005) concluded that Putnam shareholders had been harmed not only by the direct dilution from employee trading, but also from the price impact caused by sales that were forced when customers withdrew funds after the scandal was revealed. This caused the restitution amount for the Putnam settlement to be adjusted from $10 million to $93 million. Given that this upward revision was likely not anticipated at the time the settlement was negotiated, I have decided not to include it in my central results, but do include an alternative version of the results in Table 10 that uses the revised restitution amount for the Putnam settlement. Adjusting the Putnam restitution amount does lower the New York involvement coefficient, especially for restitution, but it remains statistically and economically significant. As a final robustness check in Table 11, I dilution-weight rather than equal-weight the settlements. Doing so significantly increases the estimated magnitude and precision of the constant and New York involvement coefficients, and also yields F and R^2 statistics that suggest a tighter regression fit. All of the results presented thus far in the paper are significantly stronger if observations are dilution weighted.[36]

As Hogue and Wellman (2005), Choi and Kahan (2006), and Schwarz and Potter (2006) have shown, mutual fund firms involved in regulatory cases experienced "market penalties" as well as regulatory ones. Table 12 compares the restitution and penalties imposed by regulators with the effect on firm value of business lost after the cases were announced. I follow Choi and Kahan (2006) and estimate the lost business penalty associated with a regulatory case by comparing the actual assets of a family's surviving mutual funds 12 months after announcement with what their assets would have been had their funds received the asset-weight average percentage inflow for their categories (defined using S&P objective codes). The average settlement firm suffered a lost business penalty of over 10 percent of assets; penalties were especially high for Invesco (40 percent), Strong (32 percent), Putnam (30 percent), and Janus (28 percent).

[35] In the other 19 settlements, the IDCs did not play a role in determining the settlement amount, but rather developed the plan for apportioning the settlement among affected shareholders.

[36] Two additional robustness checks were omitted from Table 11. In the first, I found that I could drop any one settlement from the sample and still obtain a statistically significant coefficient for New York's involvement. In the second, I found that results were robust to using actual dilution rather than predicted dilution from fund share turnover (this resulted in dropping four of the settlements from the sample).

Table 12. Comparison of regulatory and "market" penalties
$ millions

					Mutual fund assets, 12 months after scandal			Total cost of scandal	Ratios to dilution		
	Estimated Dilution (from Table 8)	Restitution	Penalty	Fee reduction	But for scandal	Actual	Value of lost assets	Rest + Pen + Fees + Outflows	Rest	Rest + Pen	Rest + Pen + Fees
Strong	14	70	70	35	33,365	22,717	426	601	5.0	10.0	12.5
MFS	35	175	50	125	92,278	80,178	484	834	5.1	6.5	10.1
Banc One	20	10	40	40	98,459	89,722	350	440	0.5	2.5	4.5
Alliance	64	150	100	350	83,911	78,331	223	823	2.3	3.9	9.3
Putnam	246	10	100	0	166,589	116,593	2,000	2,110	0.0	0.4	0.4
Bank of America	117	250	125	160	114,489	106,800	308	843	2.1	3.2	4.6
Janus	255	50	50	125	127,601	93,066	1,381	1,606	0.2	0.4	0.9
Federated	96	35	45	20	158,396	146,282	485	585	0.4	0.8	1.0
Invesco	150	210	115	75	21,327	12,757	343	743	1.4	2.2	2.7
Fleet	89	70	70	0	61,170	56,124	202	342	0.8	1.6	1.6
AIM	228	20	30	0	109,073	93,700	615	665	0.1	0.2	0.2
Scudder	268	102	20	86	98,276	87,262	441	649	0.4	0.5	0.8
Pilgrim	282	160	110	10	21,937	18,269	147	427	0.6	1.0	1.0
Alger	43	30	10	5	3,624	3,312	12	57	0.7	0.9	1.0
Fremont	19	2	2	0	3,377	2,929	18	22	0.1	0.2	0.2
RS	70	12	14	0	8,736	9,165	-17	9	0.2	0.4	0.4
Franklin	479	30	20	0	202,619	205,655	-121	-71	0.1	0.1	0.1
Pimco	41	2	18	0	161,115	171,940	-433	-413	0.0	0.5	0.5
Totals	2,518	1,388	989	1,031	1,566,341	1,394,800	6,862	10,269	0.6	0.9	1.4
NY involved	1,751	1,346	851	1,031	1,036,018	900,612	5,416	8,644	0.8	1.3	1.8
NY not involved	767	42	138	0	530,323	494,188	1,445	1,625	0.1	0.2	0.2

Notes: Mutual fund assets "but for" the scandal are calculated by calculating what the assets of each fund would have been had they received the asset-weighted average percentage inflows for their cat objective code) in each of the 12 months following disclosure of the SEC/State investigation. The value of lost assets is the difference between the "but for" and actual assets of surviving mutual funds, n 4% valuation multiple (Pozen, 2002; Huberman, 2006). Dates of initial investigation disclosure is taken from Table I of Hogue and Wellman (2005).

If I value their lost assets using the approximate average valuation multiple of 4 percent from mutual fund acquisitions (Pozen, 2002; Huberman, 2006), I find that total lost business penalties ($6.8 billion) were larger than the total penalties and restitution in the settlements ($2.4 billion).[37] Adding lost business effects brings the total punishment-to-harm ratio above 1.0 for 17 of the 20 settlements. This is significant, given that a ratio of at least *1/p*, where *p <= 1* is the probability within the statue of limitations, is likely necessary for deterrence (Becker, 1968).

That said, lost business penalties vary considerably. Firms who received prosecutorial and media attention earlier in the scandal suffered much greater market penalties for a given amount of shareholder harm, and the ratio of lost business penalty shareholder harm varies even more than restitution ratios and do not appear to be systematically related to the severity of conduct. Including outflow penalties in restitution+penalty ratio significantly raises the average ratio, but does not change one's conclusion that penalties were higher in cases involving New York.[38]

VI. Why a New York Effect?

Taken together, these results suggest that New York's involvement had a significant effect on the outcome of settlement negotiations. When New York was involved, the dilution-weighted average restitution ratio was 0.77, when it was not involved, the average was 0.07. Including penalties, the corresponding averages were 1.25 and 0.25. What can account for this large a difference?

In explaining its reluctance to compare settlement amounts to family-level estimates of dilution to the House Judiciary Committee as quoted above, the SEC cited three reasons why estimates of dilution may differ: 1) the "method of estimation employed", 2) "the time period", and 3) "the specific trades during that period that are analyzed." While the robustness checks in Table 11 suggest that the overall method chosen does not significantly influence conclusions about the

[37] Choi and Kahan (2006) value lost assets at one year's expense ratio revenue and conclude that they were smaller than the regulatory penalties. For data availability reasons, Hogue and Wellman (2005), Choi and Kahan (2006), Schwarz and Potter (2006), and I are unable to calculate the "inflow penalty" suffered by the firms in their other business lines (e.g., pension or hedge fund management, brokerage). Anecdotal evidence from news reports suggests that business losses in these other product lines were also significant.

[38] Specifically, when I regress the ratio of total scandal cost (restitution, penalties, fee reductions, and outflows) to dilution reported in the last column of Table 12 on the New York involvement dummy and include the two controls for egregious conduct, I get a coefficient of 12.2 (robust standard error 5.8; p-value 0.05).

effect of New York's involvement, Table 3 illustrates an important interaction effect between time period and method.

As discussed in the previous section, the primary difference between the consulting and academic methods is that the former includes the effect of short-term trading on portfolio composition. In practice, this is often done by assuming that fund managers hold all investments from short-term traders are in cash and thus short-term trading increases the cash holdings of the fund. Ex post, holding more cash turns out to be a benefit in down markets, and thus the version of the consulting method that assumes managers target minimum cash holdings tends to yield lower estimates of dilution in down markets.

Thus one plausible partial explanation for the settlement outcome differences would be if the SEC focused attention on the downward market period of 2001-3 and favored the "min cash target" consulting methodology when setting settlement amounts. A second plausible explanation would be if New York and the SEC differed in the proportion of trades they held fund companies responsible for. Many of the SEC releases that accompanied settlements refer to the specific amount of dilution arising from favored arbitrageurs, and none that I have seen refer to the overall amount of dilution in the funds arising from a failure to fair value price. If New York considers the overall level of dilution when negotiating settlement amounts, while the SEC considered only the dilution from the favored subset of arbitrageurs, this could also explain the very different outcomes. In interviews with regulators, I have been told that both explanations played a role in why New York's involvement led to different results.

VII. Discussion

The sample size in this paper is admittedly small, with only 20 settlements and only four that do not involve New York. That said, the effect of prosecutor identity is large enough that its sign can still be convincingly established, or at least as convincingly as one could hope for from such a small sample. The difference between the New York and non-New York settlements is of interest for multiple reasons. At the most basic level, by comparing settlement amounts with shareholder dilution, they complement the GAO (2005) review of the settlement negotiation process and answer an outstanding question from the accompanying Congressional hearing.

Beyond this, the apparent effect of New York's involvement is jointly informative about the scope for prosecutorial discretion in these cases and about how New York and the SEC exercised it. Given what we already know anecdotally about New York's aggressiveness, particularly under Eliot Spitzer, that New York was more aggressive or more effectively aggressive than the SEC is perhaps not surprising. But it is arguably more surprising that New York's extra

aggressiveness affected outcomes. Beyond a certain point, prosecutorial aggressiveness should not affect outcomes, since the constraint of what can be obtained in court will bind. The large estimated effect of New York's involvement implies that the SEC, when negotiating without New York, did not reach this level of aggressiveness, even in cases in which some had worried that the political environment would lead them to be overly aggressive.

The apparent difference in regulatory aggressiveness is consistent with the incentives created by regulators' career concerns. As Reinganum (1981) argues, the optimality of allowing prosecutors discretion depends crucially on how their private incentives differ from maximizing social welfare. Both those who believe Eliot Spitzer's political career concerns led him to be overly aggressive and those who worry that the SEC staff's career concerns lead them to take pro-industry positions can find some support for limits on discretion in these results. A simple policy reform that might limit both forms of discretion would be to make mandatory the calculation and public reporting of restitution ratios, perhaps using multiple methodologies if necessary. This reform would have the side benefit of allowing future analysis of these issues with a much larger sample size. This disclosure would not be without precedent. For instance, when announcing the settlement of five Canadian market timing cases, the Ontario Securities Commission reported calculations of profits earned by fund companies and arbitrageurs, so that citizens could evaluate the relative size of the settlements. While the average restitution ratios were lower than in the U.S. settlements, the ratios varied less, perhaps because transparency restrained prosecutorial discretion.[39]

Support in these results could also be found for the idea (Stigler, 1971; Macey, 1992) that single-industry regulators (like the SEC) are less aggressive than those who regulate multiple industries (like state attorneys general), as they have staff who are more dependent on one industry for their next job. In contrast, the results do not seem consistent with the idea that the embarrassment of having under-regulated mutual fund valuation and trading before 2003 led the SEC to be overly aggressive in response. In fact, the low restitution ratios obtained by the SEC when negotiating without New York are all the more surprising given the visibility of these cases and the fact that in these cases, unlike most others handled by the SEC, it is at least possible to assess the adequacy of settlements with publicly available data. If the data become available to researchers, further work

[39] The ratio of restitution to arbitrageur profit ranged from 0.41 to 0.61 in the five Canadian settlements. For reasons discussed above, in the largely down market periods covered by the settlements, arbitrageur profit is likely to underestimate shareholder harm. Furthermore, only the profits of specific arbitrageurs were included in these calculations; the total harm to shareholders from stale price arbitrage was not disclosed.

on prosecutorial discretion in lower profile cases would provide a useful complement to these results.

References

Abrams, Robert and Joel Cohen. 2004. "Explaining Eliot Spitzer: Eliot Spitzer, The New York State Attorney General Understands How to Use Power," *Barron's*, March 22, 52.

Angrist, Josh, and Alan Kreuger. 2001. "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," 15 *Journal of Economic Perspectives*, 69–85.

Becker, Gary. 1968. "Crime and Punishment: An Economic Approach," 76 *Journal of Political Economy*, 169-217.

Bhargava, Rahul, Ann Bose, and David Dubofsky. 1998. "Exploiting International Stock Market Correlations with Open-End International Mutual Funds," 25 *Journal of Business, Finance, and Accounting*, 765-773.

Blume, Marshall and Donald Keim. 2006. "'Stale' or 'Sticky' -- What Motivates Late Trading and Market Timing in Mutual Funds?" Wharton mimeo.

Boudoukh, Jacob, Matthew P. Richardson, Marti Subrahmanyam, and Robert F. Whitelaw. 2002. "Stale Prices and Strategies For Trading Mutual Funds," 58 *Financial Analysts Journal*, 53-71.

Chalmers, John, Roger Edelen, and Gregory Kadlec. 2001. "On the Perils of Security Pricing by Financial Intermediaries: The Wildcard Option in Transacting Mutual Fund Shares," 56 *Journal of Finance*, 2209-36.

Choi, Steven and Marcel Kahan. 2006. "The Market Penalty from Mutual Fund Scandals," NYU working paper.

Ciccotello, Conrad and Jason Greene. 2006. "Mutual Fund Dilution from Market Timing Trades," 4 *Journal of Investment Management*.

Coffee, John. 2003. "The Mutual Fund Scandals: What Should the SEC Do?" Columbia Law School Center on Corporate Governance mimeo (available at http://www.law.columbia.edu/center_program/corp_gov/corp_govwrit#9055).

Cutler, Steven. 2003. *Testimony Concerning Recent Activity to Combat Misconduct Relating to Mutual Funds Before the Senate Subcommittee on Financial Management, the Budget, and International Security, Committee on Governmental Affairs*, (November 3)

Dechert, LLP. 2004. *New York State's Martin Act: A Primer* (available at http://www.dechert.com/library/FS_2004-04.pdf).

Freeman, John. 2004. *Hidden Cost and Fees of Mutual Fund Fees. Testimony Before the Senate Governmental Affairs Subcommittee on Financial Management, the Budget, and International Security*, (January 27).

Goetzmann, William, Zoran Ivkovic, and Geert Rouwenhorst. 2001. "Day Trading International Mutual Funds: Evidence and Policy Solutions," 36 *Journal of Financial and Quantitative Analysis*, 287-309.

Government Accountability Office. 2005. *Mutual Fund Trading Abuses: SEC Consistently Applied Procedures in Setting Penalties, but Could Strengthen Certain Internal Controls*, GAO-05-385 (May 16).

Greene, Jason and Charles Hodges. 2002. "The Dilution Impact of Daily Fund Flows on Open-End Mutual Funds," 65 *Journal of Financial Economics*, 131-58.

Hamermesh, Lawrence. 2006. *In the Matter of Columbia Management Advisors, Inc. and Columbia Fund Distributors, Inc.: Proposed Plan of Distribution*. (Available at http://www.sec.gov/litigation/admin/2006/34-54175-pdp.pdf).

Helland, Eric. 2001. "Prosecutorial Discretion at the EPA: Some Evidence on Litigation Strategy,' 19 *Journal of Regulatory Economics*, 271-294.

Hortacsu, Ali and Chad Syverson. 2003. "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of the S&P 500 Index Funds," NBER Working Paper No. 9728.

Hogue, Todd and Jay Wellman. 2005. "Fallout from the Mutual Fund Trading Scandal," *Journal of Business Ethics*, 62, 129-139.

Huberman, Gur. 2006. "Is the Price of Money Managers Too Low?" Columbia Business School mimeo.

James, Christopher. 2007. *In the Matter of Janus Capital Management, LLC: Proposed Plan of Distribution* (Available at http://www.sec.gov/litigation/admin/2007/34-55835-pdp.pdf).

Kadlec, Gregory and Douglas Patterson. 1999. "A Transactions Data Analysis of Non-Synchronous Trading," *Review of Financial Studies* 12, 609-630.

Kondo, Jiro. "Self-Regulation and Enforcement in Financial Markets: Evidence from Investor-Broker Disputes at the NASD," Northwestern University mimeo.

Lo, Andrew and Craig MacKinlay. 1990. "An Econometrics Analysis of Non-Synchronous Trading," *Journal of Econometrics* 45, 181-211.

Macey, Jonathan R. 1992. "Organizational Design and Political Control of Administrative Agencies," *Journal of Law, Economics, and Organization* 8, 93-110.

Macey, Jonathan R. 2005. "Positive Political Theory and Federal Usurpation of the Regulation of Corporate Governance: The Coming Preemption of the Martin Act," *Notre Dame Law Review* 80, 951-974.

Mahoney, Paul. 2004. "Manager-Investor Conflicts in Mutual Funds," 18 *Journal of Economic Perspectives*, 161-182.

Massa, Massimo, Jonathan Reuter, and Eric Zitzewitz. 2009. "When Should Firms Share Credit with Employees? Evidence from Anonymously Managed Mutual Funds," *Journal of Financial Economics*, forthcoming.

Nagar, A. L. 1959. "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations," *Econometrica* 27(4), 575-595.

O'Brien, Justin. 2005. "The Politics of Enforcement: Eliot Spitzer, State-Federal Relations, and the Redesign of Financial Regulation," *Publius* 35(3), 449.

Paternoster, Raymond. 1984. "Prosecutorial Discretion in Requesting the Death Penalty: A Case of Victim-Based Racial Discrimination," 18 *Law & Society Review*, 437-478.

Pozen, Robert. 2002. *The Mutual Fund Business, 2nd Edition*. Houghton Mifflin, Boston.

Radelet, Michael L. and Glenn L. Pierce. 1985. "Race and Prosecutorial Discretion in Homicide Cases," 19 *Law & Society Review*, 587-622

Reinganum, Jennifer. 1981. "Plea Bargaining and Prosecutorial Discretion," 78 *American Economic Review*, 713-728.

Romano, Roberta. 2001. "The Need for Competition in International Securities Regulation," *Theoretical Inquiries in Law* 2, 387.

Schmidt, Janell and Steury, Ellen H. 1989. "Prosecutorial Discretion in Filing Charges in Domestic Violence Cases, 27 *Criminology*, 487-510.

Schwarz, Christopher and Mark Potter. 2006. "The Mutual Fund Scandal and Investor Response," Babson College working paper.

Stigler, George. 1971. "The Regulation of Industry," 2 *Bell Journal of Economics and Management Science*, 3-21.

Tufano, Peter. 2005. *Report of Peter Tufano, Independent Assessment Consultant in the Matter of Putnam Investment Management, LLC, Administrative Proceeding File No. 3-11317*. (Available at http://www.sec.state.ma.us/sct/sctpdf/ptnreportemp.pdf).

United States House of Representatives. 2005. *Mutual Fund Trading Abuses: Hearing Before the Subcommittee on Commercial and Administrative Law of the Committee on the Judiciary* (available at http://judiciary.house.gov/media/pdfs/printers/109th/21853.pdf, last assessed April 30, 2007).

Woodward, Susan. 2001. "Regulatory Capture at the Securities and Exchange Commission," in *Restructuring Regulation and Financial Institutions*, ed. James R. Barth, R. Dan Brumbaugh Jr. and Glenn Yago, Milken Institute, Santa Monica, CA.

Zitzewitz, Eric. 2003. "Who Cares About Shareholders? Arbitrage-proofing Mutual Funds," 19 *Journal of Law, Economics, and Organization*, 245-280.

Zitzewitz, Eric. 2005a. *Testimony of Eric W. Zitzewitz, Assistant Professor of Economics, Stanford Graduate School of Business, before the Subcommittee on Commercial and Administrative Law, United States House of Representatives, On "Mutual Fund Trading Abuses,*, (June 7).

Zitzewitz, Eric. 2005b. "Update on the Pricing and Dilution of Mutual Funds," Stanford mimeo.

Zitzewitz, Eric. 2006. "How Widespread Was Late Trading in Mutual Funds?" 95 *American Economic Review (Papers & Proceedings)*, 284-289.