Finding Bernie Madoff: Detecting Fraud by Investment Managers

Stephen G. Dimmock and William C. Gerken*

January 16, 2011

Abstract

Using a panel of mandatory SEC disclosure filings we test the predictability of investment fraud. We find that past regulatory and legal violations, conflicts of interest, and monitoring have significant power to predict fraud. Avoiding the 5% of firms with the highest ex ante predicted fraud risk allows investors to avoid 29% of investment frauds and over 40% of the total dollar losses from fraud. Even though our predictions are based on publicly available information, we do not find evidence that investors are compensated for fraud risk through superior performance or lower fees. The results suggest that the currently required disclosures contain relevant information but this information is not fully utilized by investors. We suggest changes in SEC disclosure policy that increase investors' ability to detect fraud at minimal cost.

JEL Classifications: G2, G20, G28, K2, K22

Keywords: Fraud, Investment Fraud, Operational Risk, Ponzi Scheme, SEC, Disclosure, Form ADV, Hedge Funds, Hedge Fund Fraud

---

*Dimmock (dimmock@ntu.edu.sg) is at Nanyang Technological University, and Gerken (will.gerken@auburn.edu) is at Auburn University. We are grateful for comments from: James Angel, Devraj Basu, Ranadeb Chaudhuri, Elroy Dimson, Will Goetzmann, Philip Hamill, Chuan-Yang Hwang, Dusan Isakov, Zoran Ivković, Jun-Koo Kang, Ayla Kayhan, Naveen Khanna, Clive Lennox, Jay Patel, Josh Pollet, Melvyn Teo, Fan Yu, Lei Zhang, seminar participants at Auburn and Nanyang Technological Universities, Singapore Monetary Authority, 2010 Asian Financial Management Association, Current Topics in Securities Regulation Conference at Notre Dame, One-Day Conference on Professional Asset Management at Erasmus University, 2010 European Financial Association, and Fourth Singapore International Conference on Finance, 2010 CRSP Forum, 2010 Financial Management Association, and the Emerging Scholars in Banking and Finance conference. We are especially grateful to Scott Weisbenner for numerous helpful discussions. We acknowledge the financial support of Networks Financial Institute at Indiana State University and its award for best paper in Financial Services Regulatory Reform for 2010-11.

On December 11, 2008 Bernie Madoff and his firm, Bernard L. Madoff Investment Securities LLC, were charged with securities fraud for an $18 billion Ponzi scheme.[1] In the wake of this scandal both law makers and the general public have exerted significant pressure to alter the regulation of investment advisors. In response, the U.S. Securities and Exchange Commission (SEC) has proposed substantial regulatory changes.[2] The pressure to act quickly, however, has resulted in proposed changes that are tailored to the specific characteristics of the Madoff fraud, and there does not appear to have been a comprehensive evaluation of the existing regulatory system.

Historically, disclosure has been the basis of U.S. investor protection laws. The SEC's report on Post-Madoff Reforms, however, scarcely mentions disclosure and instead focuses on expanding the role of public enforcement of securities laws, such as through SEC examinations of investment advisors. The implicit assumption behind this shift in emphasis is that the existing system of mandatory disclosure is insufficient.

The primary federal law mandating disclosure is the Investment Advisers Act of 1940. This law requires investment advisors to file Form ADV, which contains information on conflicts of interest as well as past regulatory and legal violations. In this paper, we use a panel of all Form ADVs filed by registered investment advisors from August 2001 through July 2006. This dataset includes 13,853 investment advisors who provide advice to more than 20 million clients and have discretionary control of more than $32 trillion in assets. The sample includes all mutual fund advisors, nearly all institutional investment fund advisors, and a large number of hedge fund advisors.

To obtain data on investment frauds, we search all SEC litigation actions and administrative proceedings from August 2001 through July 2010 and identify all cases in which investment advisors defraud their clients. From these filings we collect information on the

[1] http://www.sec.gov/news/press/2008/2008-293.htm
[2] http://www.sec.gov/spotlight/secpostmadoffreforms.htm

type of fraud, time-span, size of investors' losses, person or persons involved, and detection date.

We then test whether the information investment advisors disclose in their Form ADV filings can be used to predict fraud. If the required disclosures are useful for predicting fraud, this would provide evidence that regulators require disclosure of relevant information. If the required disclosures are not useful for predicting fraud this could be interpreted in two ways: 1) The required disclosures are not useful or 2) The required disclosures have a strong deterrence effect, which offsets any relation between disclosed information and the commission of fraud.

Using the Form ADV data, we show that disclosures related to past regulatory violations, conflicts of interest, and monitoring all significantly predict fraud. Avoiding the 5% of firms with the highest ex ante fraud risk allows an investor to avoid 29% of frauds and over 40% of the dollar losses from fraud. This predictability is robust in out-of-sample tests.

The purpose of these regressions is to predict fraud, and many of our independent variables are jointly determined with the decision to commit fraud. Thus we do not make causal inferences based on these regressions; even if certain practices predict fraud, the regressions cannot be interpreted to imply that there would be any benefits from banning these practices.

Our study also examines the considerable heterogeneity among frauds. Investor losses are much higher in theft cases than in cases in which a firm merely misrepresents past performance. Similarly, fraud by a rogue employee is generally less damaging than fraud orchestrated by the executives of a firm. To ensure our results are not driven only by less economically damaging forms of fraud, we test the predictability of different types of fraud. While frauds involving theft are relatively more difficult to predict, avoiding the 5% of firms with the highest fraud risk allows an investor to avoid 27.4% of thefts. We find similar results for firm-wide frauds (as opposed to fraud by rogue employees).

The predictability of fraud raises the question: why do investors allocate money to high fraud risk firms? One possibility is that investors are compensated for fraud risk through superior performance or lower fees. To test this, we merge our firm level data with fund level data for the subset of firms that report to the TASS hedge fund, CRSP mutual fund, and/or PSN Informa databases. In all three samples of funds, we find no evidence that investors are compensated for fraud risk.

Given the surprising result that fraud risk is both predictable and uncompensated, we next examine the barriers and costs to implementing the predictive methods developed in this paper. We show that the method through which the SEC provided Form ADV data to investors moderately reduces the predictive value of the disclosures and also increases the costs of processing the data. We discuss simple (and virtually costless) changes to the historical disclosure format that substantially improve investors' ability to use the disclosed data to predict fraud.

1 Related Research

To the best of our knowledge, Bollen and Pool (2010) and Zitzewitz (2006) are the only other papers that test whether it is possible to detect fraud by investment managers. Bollen and Pool (2010) build on several earlier studies that test whether hedge funds manipulate reported returns (e.g. Bollen and Pool (2008), Bollen and Pool (2009), and Straumann (2008)), and find that suspicious return patterns have some ability to predict future fraud charges. Zitzewitz (2006) shows that daily flow data provide information about mutual fund late trading. While similar in spirit to our paper, these studies use return and volume data rather than firm characteristics. An advantage of using firm characteristics is that we are able to predict fraud prior to its initiation, whereas methods based on returns and volume are only able to detect fraud after it has occurred.

Brown, Goetzmann, Liang, and Schwarz (2008, 2009) examine the topic of operational risk using a cross-section of Form ADV filings by hedge fund management firms. They define "problem" funds as those managed by a firm that reports any prior legal or regulatory violations committed by either the firm itself or any affiliated firm (defined as firms under common control). Brown et al. then test whether current firm characteristics are associated with past problems. Because historical Form ADV data are not publicly available, the authors create a measure of operational risk based on the correlations between Form ADV data and historically available hedge fund data, and then test if this measure predicts hedge fund death, flows, and returns.

Although we also use Form ADV data, our work differs from Brown, Goetzmann, Liang, and Schwarz (2008, 2009) in several ways. First, we focus on predicting fraud rather than the very broad definition of operational risk used by Brown et al. Indeed, we find that of the 126 investment advisors Brown et al. identify as having operational risk problems, only 6 have an incident of fraud. Second, their measure of operational risk includes violations by affiliated firms such as broker-dealers (e.g. hedge funds managed by the firm Wall Street Access Management, LLC would be labeled problem funds because of minor trading violation committed by an affiliated brokerage firm). These differences are empirically important; when we replicate the ω-score of Brown et al. we find it has an insignificant negative relation with future fraud. Third, we use historical Form ADV filings to make ex ante predictions of fraud. Finally, our sample of investment managers is much more comprehensive as it includes mutual funds and pension funds, in addition to hedge funds.

2 Mandatory Disclosure and Fraud

What is the optimal level of fraud? Although the instinctive response is zero, the cost of completely eliminating fraud is almost certainly higher than the benefits. Instead, it is socially optimal to allocate resources to fraud elimination only until the marginal benefit from a further reduction in fraud is equal to the marginal cost of increased enforcement (see Becker (1968)). The fact that the SEC's report on Post-Madoff Reforms proposes changes to the current regulatory system implies that either the perceived marginal benefit of reducing fraud has increased or the perceived marginal cost has changed. Possibly the Madoff case has increased the estimated extent of fraud, implying higher marginal benefits from fraud reduction. The proposed changes, however, include not only an increase in the *level* of expenditures on fraud reduction but also a change in the *composition* of expenditures - towards greater public enforcement of securities laws. This change in the composition of expenditures suggests a decrease in the perceived marginal benefits of disclosure relative to the marginal benefits of direct enforcement.

Ideally we could directly test whether the marginal cost of disclosures is equal to the marginal benefits of fraud reduction. This is not, however, an empirically feasible strategy. First, we have no way of separating marginal versus average costs and benefits. Second, the benefits of disclosure include deterrence, and deterred frauds are inherently unobservable. Instead, we test the equilibrium implications of an effective disclosure system.

Suppose the disclosed information in Form ADV is not useful for predicting fraud. This might indicate that the data are worthless. Alternatively, a lack of predictability could be consistent with equilibrium for two closely related reasons. First, the deterrent effect of disclosure could be sufficient to eliminate predictability. For example, suppose that the use of soft dollars provides a mechanism for committing fraud. Then, in the absence of disclosure, we would expect fraudulent investment advisors to use soft dollars. Disclosing soft dollar use, however, could increase investor monitoring, and the resulting increase in the probability of detection could offset any benefit a fraudulent investment manager gains from using soft dollars. Second, disclosure is not the only mechanism investors use to reduce fraud risk. Investors can respond to disclosed information by increasing direct monitoring or by demanding audits. By changing the level of investor monitoring, disclosure could eliminate the ability of the disclosed information to predict fraud.

Suppose, on the other hand, that the disclosed information is useful for predicting fraud. Provided that investors are compensated for higher fraud risk in some way, this result is consistent with equilibrium. Depending on their tastes, some investors would accept a relatively high level of fraud risk in return for lower fees or higher performance while other investors would choose low fraud risk and pay higher fees or accept worse performance. If this is the case, higher regulation might reduce the rate of fraud but also decrease aggregate investor welfare, as investors who are comfortable bearing fraud risk could no longer earn compensation for bearing this risk (see Karpoff and Lott (1993) for a detailed discussion).

Even if the disclosed data are useful for predicting fraud, and fraud risk is not compensated, this does not necessarily imply that market participants are irrational. The costs of obtaining the data and estimating fraud risk may well exceed the perceived benefits for many investors. This problem is exacerbated by the fact that investors are atomistic: even if the aggregate benefit of processing the disclosed information is greater than the cost for a single investor, the benefit to any individual investor may be insufficient. Monitoring is in part a public good, because all investors benefit from the disincentive for fraud created by the monitoring of one investor, as a result monitoring is subject to a free-rider problem. Also, because it is not possible to sell short many types of investment funds, investors who correctly identify fraud risk can only benefit by avoiding fraud risk which limits the incentive to measure fraud risk.

3 Legal Background

The Investment Advisers Act of 1940 requires SEC registration for all investment advisors with more than $25 million in assets under management and 15 or more U.S. based clients.

The Act defines investment advisors as any entity that receives compensation for managing portfolios of securities for clients or provides advice regarding individual securities (e.g. pension fund and mutual fund advisors). Registered investment advisors must file Form ADV to disclose past regulatory violations and potential conflicts of interest. The Investment Advisers Act also prohibits fraud by investment advisors.

Section 203(b)(3) of the Investment Advisers Act exempts investment advisors that during the preceding 12 months had fewer than 15 U.S. clients,[3] do not advise investment companies registered under the Investment Company Act of 1940, nor "hold themselves out to the public" as investment advisors. Some hedge funds use this exemption to avoid registration. A rule passed by the SEC required hedge fund managers to register by February 1, 2006, but this rule was reversed in June of that year. Despite these exemptions, many firms that manage hedge funds were registered prior to 2006; either because they also managed other investment portfolios, had more than 15 clients, or voluntarily registered.

The other major law governing investment managers is the Investment Company Act of 1940, which covers mutual funds and other investment companies targeting retail investors. The Investment Company Act provides additional investor protection and requires numerous additional disclosure filings. The Investment Advisers Act covers a related, but broader, set of investment firms. For example, the investment advisor, Fidelity Management and Research Company (covered by the Investment Advisers Act), advises the Fidelity family of mutual funds (covered by the Investment Company Act).

[3] Hedge fund avoid violating the 15 client rule by forming partnerships of fewer than 100 investors and then counting the partnership as a single client. This does not work for large hedge fund management companies. For example, Madoff had to file Form ADV as he had more than 15 clients.

4 Data

4.1 Investment Fraud

We combine two data sources in this study: 1) data on investment fraud and 2) investment advisor characteristics disclosed in Form ADV. To obtain data on investment frauds we identify all SEC administrative proceedings and litigation releases[4] that contain the word "fraud" and the phrase "investment advisor" (or "investment adviser") from August 2001 through July 2010. From these filings we identify all legal actions that involve violations of the antifraud provisions in the Investment Advisers Act. The main dependent variable includes only cases of fraud that directly harm the firm's investment clients. We do not include insider trading, short sales violations, crimes by the brokerage division of the firm, or other activities, unless these crimes cause direct losses to the firms' investment clients (e.g. we exclude cases where investment advisors generate profits for their clients through insider trading). Even when fraud is detected by an agency other than the SEC, the SEC launches an administrative action which we observe.

Because we wish to test whether prior fraud predicts future fraud, we also collect information on all frauds committed by investment advisors and broker-dealers from January 1995 through July 2006 using the same data sources. Frauds committed by affiliated firms are matched to registered investment advisors using the affiliated firm identifiers listed in Schedule D of Form ADV (affiliated firms are firms under common control i.e. with common ownership or executives). Prior fraud, which is used as an independent variable, differs from our main dependent variable in that it includes frauds that harmed investment clients as well as other forms of fraud.

Many frauds occur over several years and are associated with multiple legal actions. For

[4] See `http://www.sec.gov/litigation/admin.shtml` and `http://www.sec.gov/litigation/litreleases.shtml`

example, Figure 1 shows the time line of one fraud in our sample. The fraud was initiated in September 2002 by K.W. Brown & Company, an investment advisor which traded securities on behalf of clients and also for its own proprietary account. K.W. Brown would purchase securities but delay assigning trades to specific accounts until the return on the trade was clear. At that point, the firm would allocate profitable trades to its proprietary account and unprofitable trades to clients. The SEC uncovered this fraud in March 2003 during a routine examination. In June 2003, the SEC notified the firm of the problems identified and requested further information. The firm continued defrauding clients for an additional nine months, until March 2004. Criminal charges were filed against the firm and its key employees in April 2005, resulting in convictions in December 2007. In January 2008, the SEC filed an administrative proceeding to bar Kevin W. Brown, his wife, and one other employee from the securities industry. The employees were barred from the industry in February 2008 and the firm was deregistered in June 2008.

Because fraud cases with extended time lines and multiple legal actions are common, we aggregate all legal and regulatory actions associated with a single underlying fraud into a single record. Since our goal is to predict fraud, not the detection of fraud, we use information in the legal filings to identify the time periods in which the fraud actually occurred. For example, in the case of K.W. Brown & Co. we define the fraud as occurring from September 2002 until the fraud ends in March 2004. Thus we use information from K.W. Brown's August 2002 Form ADV filing to predict the initiation of the fraud in September 2002, and in many of our specifications, we use information from the August 2003 Form ADV filing to predict the continuation of the fraud into 2004. For the remaining years in sample, we classify K.W. Brown & Co. as a clean firm. By predicting the occurrence of fraud, rather than detection, we avoid potential biases caused by a correlation between detection and time variation in the predictive variables.

We collect information on all investment frauds, including those committed by firms which do not file Form ADV. Panel A of Table 1 shows that slightly over half of investment frauds are committed by SEC registered investment advisors, and the remainder are committed by unregistered investment advisors.[5] Although registered investment advisors commit slightly over half of investment frauds, they are responsible for the overwhelming majority of the dollar value of fraud. We also divide the fraud cases into firm-wide frauds, committed with the knowledge of the firms' executive officers, and fraud by rogue individuals who evade their firms' internal controls. The vast majority of frauds are firm-wide.

Panel B of Table 1 summarizes the types of fraud in the sample. Although many frauds involve offenses in multiple classifications, we assign each fraud into a single classification based on its most serious aspect. For example, all frauds classified as Ponzi schemes also involve misrepresentations about the firm's investment activities. However, because the Ponzi scheme is far more serious, these frauds are classified as Ponzi schemes and not Misrepresentation. Ponzi schemes combine theft with payments of stolen money to early investors so as to evade detection and lure in new investors. Direct Theft also involves stealing money, but without the payments to early investors that characterizes a Ponzi scheme. Self Dealing includes all types of fraud in which an advisor illegally profits from their clients' trades (e.g. front running clients' trades or ex post allocation of trades). Overstating Assets occurs when an investment manager overstates returns or asset values, and charges unwarranted fees based on these inflated values. Mutual Fund Late Trading includes the well publicized cases in which mutual funds allowed certain investors to place trades after closing. Misrepresentation occurs when an investment advisor lies to attract new investors (e.g. misrepresenting assets under management, or past regulatory violations). Direct Theft is the most common type of fraud, followed closely by Self Dealing.

Panel B of Table 1 summarizes the dollar value of frauds. This information is missing for

[5]Although recent regulatory proposals focus on unregistered hedge funds, only 26.9% of frauds by unregistered advisors are hedge fund frauds.

some frauds, and when available, the reported amount is generally a lower bound. Determining the exact value of fraud is difficult as a key aspect of many frauds is the falsification of records. In many cases, the reported losses include only the amount definitively proven to have occurred due to fraud. Fraud size, as a percentage of assets invested, is largest for Ponzi and Direct Theft, and smallest for Overstating Assets and Misrepresentation.

Panel B of Table 1 also summarizes the duration of frauds, defined as the time period from the initiation of the fraud until the first relevant legal filing by the SEC. The median fraud persists for five and a half years before detection. Ponzi schemes have the longest duration, suggesting they are particularly difficult to detect. The relatively high maximum values in the final column reflect the fact that some frauds were initiated prior to the beginning of our sample.

4.2 Investment Management Firm Characteristics and Disclosures

The SEC requires all registered investment advisors with at least 15 U.S. clients and more than $25 million in assets under management to file Form ADV annually or upon material changes. Form ADV contains 12 items and three schedules. Items 1 through 6 contain descriptive information about the firms' legal structure and its operations. Items 7 and 8 require disclosure of certain conflicts of interest. Item 9 requires disclosure regarding the custody of client assets. Item 10 requires disclosure of control persons. Item 11 requires disclosure of past legal and regulatory violations. Item 12 reports information about small businesses. Schedules A and B require the firm to disclose direct and indirect owners. Schedule D requires disclosure of affiliations with other financial firms.

The data include a complete set of all Form ADV filings from August 2001 through July 2006, including initial filings, amendments, schedules, and the filings of defunct firms. The SEC does not make these historical filings publicly available, and these data have not been examined by other researchers. To create an annual panel dataset, we select each firm's current filing as of August 1^{st}.[6] Our sample includes 53,994 firm-year observations representing 13,853 unique investment management firms.

We match the fraud data to the Form ADV sample using firms' full legal names. Form ADV contains each firm's full legal name, as do the administrative actions and litigation releases used to identify frauds. When multiple firms have very similar names we verify the match based on location and other information contained in both samples. The administrative proceeding and litigation release documents state whether an investment advisor is registered. We are unable to match 13 of the fraud filings related to registered investment advisors, because in these cases the firm ceased filing Form ADV before our sample began.

4.2.1 Descriptive Information

Panel A of Table 2 summarizes descriptive information about firms' investment advisory business. There is large variation in assets under management (AUM). While median AUM is $88 million, the mean is greater than $2.5 billion. Average account size is also highly skewed, with a mean substantially larger than the 75^{th} percentile.

In addition to descriptive information on their advisory businesses, Form ADV requires firms to disclose a large amount of information. The disclosure variables are summarized in Panel B of Table 2 (see Appendix Table 1 for detailed definitions of the variables). Column one shows pooled averages across all firm-year observations. Column two shows summary statistics for firm-year observations in which there is not an ongoing fraud (clean firms). Column three summarizes firm-year observations in which fraud is ongoing (fraud firms). Column three also shows that all of the variables except for Chief Compliance Officer are significantly different between clean and fraud firms.

[6]We choose August 1^{st} to maximize the number of annual observations since our dataset of ADV filings ends July 31^{st}, 2006.

4.2.2 Disclosure of Prior Legal and Regulatory Violations

Item 11 of Form ADV, titled Disclosure Information, requires each investment advisor to disclose the disciplinary history of the firm, its employees (other than clerical employees), and affiliated firms. Firms must answer 24 questions divided into three categories: criminal disclosures, regulatory disclosures, and civil judicial disclosures. Brown, Goetzmann, Liang, and Schwarz (2009) use these disclosures to identify "problem" funds and use this as their main dependent variable. In contrast, we use these disclosures as independent variables to predict fraud. Specifically, we create two indicator variables. Past Regulatory equals one if the firm discloses any past regulatory violations, indicating sanctions by the SEC, Commodity Futures Trading Commission, or a self regulatory organizations such as the Financial Industry Regulatory Authority (FINRA). Past Civil or Criminal equals one if the firm discloses any unfavorable investment related civil judicial decisions or criminal convictions. Fraud firms are nearly three times as likely to report both types of violations.

The disclosed regulatory and legal issues cover a wide range of items and are often very minor: e.g. violations of record storage protocols or missing legal filing deadlines by a few days. Minor violations seem to be the norm rather than the exception, and should be interpreted as such; less than 4% of firms that report past violations have prior instances of fraud.

Investment advisors must disclose their own prior violations as well as prior violations by all affiliated firms (firms under common control). The responses in Item 11 do not differentiate violations attributable to the firm from violations attributable to affiliated firms. For example, one hedge fund management firm reports past regulatory violations because it is owned by a firm whose Japanese brokerage subsidiary violated Tokyo Stock Exchange rules. Another hedge fund management firm reports past criminal convictions because an executive at an affiliated broker/dealer was convicted of assault in relation to a bar fight.

Because investment advisors must disclose prior violations by both themselves and their affiliates, prior violations have a strong positive correlation with the number and size of affiliated firms. To avoid a spurious correlation between prior violations and fraud, in all specifications the dependent variable includes only frauds committed by the registered investment advisor filing Form ADV and does not include frauds committed by affiliated firms.[7]

4.2.3 Conflicts of Interest and Other Characteristics

Items 7 and 8 of Form ADV require firms to disclose conflicts of interest. We use this information to create several variables. Referral Fees is an indicator variable equal to one if the firm compensates other parties for client referrals. Referral fees are legal, provided they are disclosed to the client. However, consistent with these fees creating conflicts of interest fraud firms are more likely to pay referral fees.

Interest in Transactions is an indicator variable equal to one if the firm either trades directly with clients or has a direct financial interest in securities it recommends to its clients.[8] Transacting directly with clients is a serious conflict of interest and also provides a mechanism for fraud. Panel B of Table 2 shows that fraud firms are significantly more likely to have an interest in client transactions.

Soft Dollars is an indicator variable equal to one if the firm directs clients' trades to a brokerage with relatively high commissions, and in return, the broker supplies the advisor with research or other benefits. Since clients pay the costs but the investment advisor realizes

[7]Consider the following example for illustrative purposes. Suppose all investment advisors are identical except that half of investment advisors are affiliated with broker-dealers and half are not. Further suppose that all investment advisors have identical fraud risk. If the dependent variable includes only frauds by investment advisors, the estimated coefficient on broker-dealer affiliation will be insignificant. But if the dependent variable includes frauds by affiliated broker-dealers, the estimated coefficient on broker-dealer affiliation will be positive and significant.

[8]Formally, this variable equals one if the firm answers "Yes" to any of questions 8A1, 8A3, 8B2, or 8B3. We do not include 8A2, as it includes investment advisors investing in their own funds, which seems unlikely to increase the probability of fraud. We do not include 8B1 (Does an affiliated brokerage execute trades for brokerage clients in securities which are also purchased for investment advisory clients) as it is very highly correlated with the Broker in Firm variable.

the benefits, soft dollars are a potential conflict of interest, which may predict fraud. Panel B shows that fraud firms are more likely to accept soft dollars.

Employee Ownership summarizes the percentage of the firm owned by employees, and is calculated following the method of Dimmock, Gerken, and Marietta-Westberg (2010). Employee ownership is very common and the median firm is wholly employee owned. We include this variable as external owners potentially provide monitoring of employees, which may deter fraud.

Percent Client Agent is the percentage of the firms' clients who are agents rather than the direct beneficiaries of the funds they allocate (e.g. pension funds). At the average firm, approximately a quarter of the clients are agents. This additional layer of agency is potentially related to fraud, as agents have weaker incentives to monitor.

Broker in Firm equals one if the firm employs registered representatives of a broker-dealer. This removes one source of external oversight and provides a mechanism for certain types of fraud. Panel B shows that affiliation with a broker/dealer is associated with higher rates of fraud.

Custody is an indicator variable equal to one if the firm has possession of, or the authority to obtain possession of, clients' assets. Custody enables fraud by removing third party oversight. However, SEC Rule 206(4)-2 requires investment advisors with custody of client assets to be audited, including at least one unannounced visit per year to verify client assets. Even with this audit requirement, Custody is higher for fraud firms.

Investment Company Act is an indicator variable equal to one if the firm manages money on behalf of a fund registered under the Investment Company Act of 1940 (e.g. mutual funds). The Investment Company Act increases regulation and disallows certain conflicts of interest. However, the implicit assumption behind the Investment Company Act is that mutual fund investors are relatively unsophisticated and require greater protection, which may or may not outweigh the additional legal protections of the Investment Company Act.

All registered investment advisors must designate a chief compliance officer (CCO) responsible for ensuring compliance with SEC regulation. Often the CCO has other (potentially conflicting) roles within the firm. Dedicated CCO indicates firms in which the CCO has no other formal job title. Dedicated CCO is not significantly different between clean and fraud firms.

Hedge Funds is an indicator variable equal to one if more than 75% of the firm's clients are hedge fund clients. Slightly over 10% of the firms in our sample primarily manage hedge funds, but only 4.7% of fraud firms manage hedge funds. We include this variable for two reasons. First, hedge funds are relatively opaque which may facilitate fraud. Second, prior to 2006 some hedge fund management firms were not required to file Form ADV. This may create a sample selection bias if non-reporting is higher for fraudulent hedge funds.

4.3 Fund Level Data: Returns and Fees

Most of the empirical tests in this paper use the full sample of firms that file Form ADV. To test the relation of fraud risk with performance and fees, however, requires fund level data that are not available in Form ADV. To obtain these data, we match firms that file Form ADV to investment managers who report information to one or more of the following databases: TASS hedge fund, CRSP mutual fund, and PSN Informa. For the CRSP mutual fund and PSN Informa databases we include only equity funds in our sample. We match Form ADV firms to these databases using firm name, location, and assets under management.

Table 3 shows that we are able to match 1,511 of the firms in the TASS database (37.2%), which manage a total of 2,848 distinct funds. From TASS we obtain monthly hedge fund returns, annual management fees, and annual incentive fees.

Because investment companies that manage mutual funds must file Form ADV, we are able to match all mutual fund families in the CRSP mutual fund database. From this database we obtain monthly return data for 2,818 actively managed equity funds, as well as the annual expense ratios for these funds.

To obtain information on institutional investment management firms we use the PSN Informa database (these firms manage long-only portfolios for accredited investors, see Busse, Goyal, and Wahal (2010) for a detailed description of the industry). We are able to match 1,578 of the PSN firms (88.2%), which manage a total of 4,189 distinct portfolios. These firms manage 89.2% of the aggregate assets under management in the PSN database. In addition to monthly returns, we also obtain information on the posted annual fee charged on a $50 million account (institutional funds do not charge all clients the same fees and the reported fee information is only the listed price).

In total, 3,123 of the firms in the Form ADV sample are matched to at least one of the fund level databases, and 314 of the firms match to all three databases. Although the matched firms control the majority of assets under management in our sample, only 22.5% of the total number of firms in the Form ADV sample are matched to a return database. Relative to the entire Form ADV sample, the matched firms are larger and more complex.

5 Predicting Fraud

In this section we test whether the information from the Form ADV filings can be used to accurately predict investment fraud. These regressions are predictive and we do not make any claims regarding causality. Many of our independent variables are endogenous. Indeed, in many fraud cases the firms' executives deliberately chose organizational structures that enabled fraud. However, since our goal is prediction, rather than establishing causality, the potential endogeneity of the independent variables does not change the interpretation of the results.

A major caveat in interpreting the findings is that we only observe detected frauds. There are three separate factors affecting fraud detection: the unobservable true rate of fraud, the probability of detection given a fixed level of monitoring, and the allocation of monitoring resources. Ideally, we want to predict the true rate of fraud. If our predictive variables are correlated with either monitoring or fraud detection, however, this may bias our results. Further, there are two reasons why the predictive variables may be correlated with fraud detection and monitoring. First, any characteristic that decreases the probability of detection increases the incentive for fraud. In general, this problem biases against finding significant results, as characteristics associated with a higher rate of fraud will also be associated with a lower detection rate. Second, if the SEC or other monitors consider the difficulty of detecting fraud when allocating monitoring resources this may, or may not, be sufficient to outweigh the added difficulty in detecting fraud. These difficulties may cause our empirical results to differ from the actual relation between firm characteristics and the unobservable true rate of fraud.

We address the problem of undetected fraud in two ways. First, although our panel of independent variables ends in 2006, we collect information on all detected frauds through July 2010. From the legal filings we identify when fraud occurred and our dependent variable is the occurrence of fraud in a given year, even if the fraud is not detected until years later. Second, we test the relation between the fraud prediction variables and the duration of detected frauds. The results, in Appendix Table 2, show that none of the independent variables are statistically significant. This is suggestive that our results are not driven by the detection rate, as it seems reasonable to assume that any variable that decreases the probability of fraud detection would be associated with longer fraud duration, conditional on the fraud being detected. Unfortunately, a direct test of the relation between these variables and fraud detection is not possible. There may be certain types of fraud which are never detected; this could bias our results and fail to be detected by the fraud duration regressions.

5.1 Prediction Models

Panel A of Table 4 shows the results of probit regressions that predict investment fraud using information from Form ADV. In all columns, the dependent variable equals one if a fraud occurs during the subsequent 12 months. In the first two columns, the sample includes all firm-year observations. In the third column, the sample excludes firms with a previously disclosed fraud. In the fourth column, the sample also excludes firms whose affiliated firms have a previously disclosed fraud. In the last column, the sample excludes all firms that disclose any type of prior legal or regulatory violation, either by the firm itself or an affiliated firm, in Item 11 of Form ADV. The results are generally similar for all samples, and so for simplicity the discussion in this subsection focuses on the first two columns and we defer a discussion of the differences across samples until the next subsection. The z-scores, reported below the coefficients, are computed using standard errors clustered by firm and by year. The chi-square test at the bottom of each column show the significance of the overall model.

The first two variables are Past Regulatory and Past Civil or Criminal. The coefficients on both variables are positive and significant even when the sample excludes all firms with prior frauds. The simplest explanation of this finding is that past problems, although frequently minor, indicate either poor internal controls or unethical management. There are, however, two additional explanations. First, past violations increase the probability of an SEC examination and this may increase the probability of fraud detection. We view this explanation as unlikely, because in Appendix Table 2 we fail to find a relation between fraud duration and past violations. Second, each firm must disclose its own prior violations as well as the prior violations of its affiliated firms. Thus prior violations are strongly correlated with the size and scope of an investment advisor's affiliated businesses (i.e. financial conglomerates are much more likely to report past violations). These affiliations may increase conflicts of interest and provide the means to commit fraud. Note that to avoid a mechanical relation between prior violations and fraud, our dependent variable includes only fraud committed by the investment advisor filing Form ADV and not fraud by affiliated firms.

Past Own Fraud and Past Affiliated Fraud are indicator variables equal to one if the firm has publicly disclosed a prior fraud by itself, or by an affiliated firm, respectively. Unlike the other predictive variables, this information is not disclosed in Form ADV and instead comes from the SEC litigation and administrative proceedings filings. Past Own Fraud has a significant positive relation with subsequent frauds. On one hand, this result is unsurprising as prior frauds signal low ethical standards. On the other hand, the disclosure of fraud should reduce the incentive to commit fraud because of a higher probability of detection due to increased monitoring by investors and regulators. The significance of the coefficient suggests that any increase in monitoring is insufficient to fully offset the underlying problems in these firms. Past Affiliated Fraud does not predict future instances of fraud.

The next four variables measure several (potential) conflicts of interest between investment advisors and their clients. The first, Referral Fees, has a significant positive relation with future fraud. Fraudulent firms may be relatively more willing to pay referral fees because fraud increases the marginal profit per dollar managed.

Firms with an economic interest in client transactions are significantly more likely to commit fraud. There is an obvious conflict of interest when investment managers take the opposite side of a transaction from their clients. Not only does this signal a lack of concern for investors, it provides a mechanism for fraud. For example, Gintel Asset Management defrauded clients by purchasing stocks for its own account and then selling the securities to investors at slightly higher prices on the same day.

Soft Dollars does not significantly predict fraud. Although the use of soft dollars may indicate a conflict of interest, the majority of clean firms in the sample accept soft dollars.

As a result, disclosed soft dollar use does not predict fraud.

Firms affiliated with broker/dealers have significantly higher rates of fraud. Using an in-house brokerage removes external oversight and creates a mechanism for committing fraud. For example, the fraud perpetrated by K.W. Brown & Company, summarized in Figure 1, was only possible because the firm conducted all trades through an affiliated brokerage.

The next four variables measure the monitoring of investment advisors. The first variable, Custody, is an indicator variable equal to one if the firm has custody of clients' cash or securities. This practice is discussed at length in the SEC's Post-Madoff Reform report, which proposes greater auditing and external oversight requirements for investment advisors with custody of client assets. The regressions, however, fail to find a significant relation between custody and fraud. The second monitoring variable indicates firms that serve as investment advisors to companies registered under the Investment Company Act of 1940 (ICA). These firms are significantly more likely to commit fraud. While the ICA increases the regulatory burden on these firms, this variable also indicates unsophisticated clients who are relatively easy to exploit. The coefficients are not significant for the next two variables, Dedicated CCO and Majority Employee Owned.

The remaining three variables also measure monitoring, but are based on the characteristics of the investment advisors' clients. Although all clients have an incentive to monitor investment advisors, both the strength of that incentive and the ability to monitor, vary across clients. The results for the first client characteristic, the logarithm of average account size, show that larger investors are associated with fewer subsequent frauds. This may be a selection effect (i.e. larger investors select honest managers). Alternatively, if large investors are financially sophisticated or enjoy economies of scale in monitoring, then large investors may decrease investment advisors' incentive to commit fraud due to a higher probability of detection. Both arguments suggest that large investors are associated with a lower rate of actual fraud, rather than a lower detection rate.

The second variable measuring client characteristics is Percent Client Agent, defined as the percentage of clients who are agents rather than the direct beneficiaries of the invested funds (e.g. pension fund managers). The coefficients are positive and significant, showing that after conditioning on average investor size, firms whose clients include a high proportion of agents are significantly more likely to commit fraud. Unlike principals, agents do not bear the full cost of a fraud. As a result, they have less incentive to exert effort in selecting and monitoring investment advisors. Also, because agents do not bear the full cost of fraud they can be swayed through gifts or kickbacks. For example, an employee of LPL Financial Corp. bribed the Treasurer of the State of New Mexico in return for business.

The final variable measuring client characteristics, is an indicator variable equal to one for firms that primarily manage hedge funds. We include this variable because hedge funds are relatively non-transparent, which potentially increases the risk of fraud. Also, there is substantial interest in hedge fund fraud due to the Madoff fraud, and other recent academic papers on the operational risk of investment managers focus on hedge funds (e.g. Bollen and Pool (2010) and Brown, Goetzmann, Liang, and Schwarz (2008)). The results in Panel A of Table 4 show no relation between hedge fund management and fraud. There are, however, two caveats. First, until the end of our sample, not all hedge funds were required to register. We have estimated a cross-sectional fraud prediction model using only Form ADV data from 2006 and fraud during the period 2007-2010. The coefficient on hedge fund management was not significantly different from the full period, suggesting there is not a sample selection problem. Second, it is possible that hedge funds have higher rates of actual fraud but their non-transparency reduces the detection rate. In Appendix Table 2, however we find no relation between hedge fund management and fraud duration.

5.2 The Economic Interpretation of the Prediction Models

The probit regressions in Panel A of Table 4 show that Form ADV data have statistically significant power to predict fraud. Statistical significance, however, does not directly address whether the model would enable investors to avoid fraud. We address this issue in three ways. Panel B shows the proportion of frauds that could be predicted within sample. Panel C summarizes the out-of-sample performance of each model, using Form ADV filings in 2006 to predict frauds that occur during 2008 through 2010. Panel D shows the results from K-fold cross-validation tests, which are explained in more detail later in this section.

5.2.1 Frauds Predicted

As discussed in Greene (2002, pg. 685), for infrequent events the standard rule of predicting an event when the estimated probability is above 50% is inappropriate and in our model such a rule would predict that no frauds ever occur. Instead we examine the proportion of frauds correctly predicted at a fixed false positive rate of 5%. False positives, which occur when the model incorrectly predicts a clean firm will commit fraud in the next year, can be thought of as the opportunity cost to investors of erroneously limiting their investment opportunity set. While it seems reasonable to think that costs are asymmetric, in that failing to predict fraud is more costly than mistakenly avoiding an honest investment manager, we do not take a strong position on cost asymmetry and instead illustrate the possible tradeoffs following a similar format to Dechow, Ge, Larson, and Sloan (2010).

The results in Panel B of Table 4 show the proportion of frauds correctly predicted within sample. The columns in Panel B correspond to the columns in Panel A (e.g. the first column in Panel B shows the percentage of frauds predicted for the regression model in the first column of Panel A). In the first column, the model correctly predicts 150 of 517 frauds (29.0%) at a false positive rate of 5% (we incorrectly predict fraud in 2,673 clean firm-years).

To complement the results in Panel B of Table 4 we also create a receiver operating characteristic (ROC) curve for the prediction model in the second column of Panel A. The points on the ROC curve, shown in Figure 2, are generated nonparametrically by using each observation's predicted value from the probit model as a cutpoint, and then computing both the proportion of frauds correctly predicted and the false positives. Random prediction of fraud would result in a straight 45 degree line. The area under the ROC curve in Figure 2 is significantly greater than the null hypothesis of no predictive power. Initially, the curve rises steeply, showing that it is possible to avoid a considerable number of frauds at little cost.

In addition to the percentage of frauds that can be avoided, we are interested in the probit models' ability to identify the largest frauds. The final row in Panel B of Table 4 shows the percentage of total dollars lost to fraud that could have been successfully avoided at a false positive rate of 5%. To avoid having extreme outliers drive our results, the dollar losses to fraud are winsorized at the 99^{th} percentile, and we exclude the Madoff and Stanford frauds. For multi-year frauds, we evenly distributed losses across all years in which the fraud occurs. The first column shows that avoiding the firms with the 5% highest fraud risk would allow an investor to avoid 41.3% of the total dollar losses from fraud. Thus the regression results are not driven by the smallest frauds.

The results in Panel B are similar in all columns except the last. In this column, the sample does not include firms that report any prior legal or regulatory violations, either by the firm or its affiliates. Both the percentage of frauds predicted and the dollar value of the frauds predicted are substantially lower for this group.

5.2.2 Out-of-Sample Prediction of Frauds

Although our sample of predictive variables ends on August 1^{st} of 2006, we collected information on all investment frauds that occurred from September 1^{st}, 2007 through July

31^{st}, 2010. Since the prediction regressions include only frauds that occurred before September 1^{st}, 2007, we use the sample of subsequent frauds to conduct an out-of-sample test of the prediction models. The results in Panel C of Table 4 show that the proportion of frauds predicted out-of-sample is actually higher than in sample, although given the small number of observations this difference is not statistically significant. Also, although we use the within-sample cutoff values to assign firms, the false positive rate in the out-of-sample classifications does not increase.

5.2.3 Cross-Validation

As a further test of the predictive validity of the regression models in Table 4, we perform K-fold cross-validation tests. Each firm in the sample is randomly assigned to one of 10 groups (note that we randomly assign *firms* and not firm-year observations, to avoid overstating the out-of-sample performance due to non-independence issues). We then estimate the prediction model 10 times, excluding each randomly formed group once. Each observation in the excluded group is assigned a predicted value using the coefficients estimated from the observations in the other nine groups. The cutoff scores for fraud prediction are calculated within sample and used to classify the observations in the out-of-sample group. We repeat this process 20 times, creating a total of 200 out-of-sample groups.

The results in Panel D of Table 4 show that predictive power of the models is only slightly lower in these out-of-sample tests. For example, the specification shown in the first column correctly predicted 150 frauds within sample, compared to an average of 143.3 frauds in the K-fold tests. The minimum number of frauds predicted across the 20 repetitions of the K-fold test is 143 frauds and the maximum is 149. The standard deviation, in number of frauds predicted, is 3.64, which suggests the model is quite stable across repetitions. Overall, the results of the out-of-sample and K-fold cross-validation tests show the predictions are robust.

5.3 Initiation versus Continuance of Fraud

The dependent variable in Table 4 does not differentiate between the initiation of a new fraud versus the continuance of a preexisting fraud. Although investors and regulators can benefit from identifying both types of fraud, identifying fraud at an early stage may reduce the total harm. To test whether Form ADV data can be used to identify fraud at an early stage, in Panel A of Table 5 we estimate a multinomial probit regression and directly compare the initiation of fraud with the continuance of a fraud. All significance tests are based on standard errors clustered by firm.

In the first column, the dependent variable equals one for firms that initiate a new fraud in the subsequent year. In the second column, the dependent variable equals one for firms that continue a preexisting fraud in the subsequent year. The excluded category is clean firms. The third column shows p-values from chi-square tests of the null hypothesis that the estimated coefficients are equal in both equations. The last row in Panel A of Table 5 shows the p-value from a chi-square test of the joint hypothesis that the coefficients are equal in both equations. The test does not reject this hypothesis. Thus, although the coefficients for Referral Fees and Broker in Firm are significantly higher in the initiation of fraud equation, this significance does not persist after adjusting for multiple comparisons.

Panel B of Table 5 shows the accuracy of the within-sample classification of frauds. At a false positive rate of 5%, we are able to predict 37.9% of frauds initiated compared with 26.0% of frauds continued. This difference in classification accuracy is significant at the 5% level. This result likely reflects a selection effect, as frauds that are relatively easy to detect do not persist long enough to enter the sample of continued, preexisting frauds.

5.4 Predicting Types of Fraud

There are many different types of fraud, ranging from direct theft to misrepresenting academic credential. There are also differences in who commits fraud. Fraud can be firm-wide, orchestrated by senior executives, or it can involve only a rogue employee. Including all frauds in a single dependent variable may reduce predictive power, as different types of fraud may be associated with different firm characteristics (e.g. theft versus mutual fund late trading).

In Panel A of Table 6 we test the predictability of different types of fraud. The first three columns show the results of a multinomial probit regression in which the dependent variables equal one for firms that commit: theft, fraudulent misrepresentation, or allow mutual fund late trading. Theft includes Ponzi schemes, direct theft, self dealing, and overstating asset values. The fourth and fifth columns show the results of a separate multinomial probit regression in which the dependent variables equal one for firms that commit: firm-wide fraud or fraud by rogue employees. In both regressions the excluded category is no fraud. All significance tests are based on standard errors clustered by firm.

In many ways, Theft is the most interesting type of fraud and the most damaging to investors. The results show that Past Regulatory violations continue to have strong predictive power. We also find that theft is more likely at firms that pay referral fees, whose clients are smaller, and whose clients are primarily agents. Panel B shows that, at a false positive rate of 5%, we are able to correctly predict 27.4% of Thefts. Although this prediction rate is slightly lower than for the overall sample, this is still a sizeable proportion of thefts and suggests that the results in Table 4 are not driven entirely by other types of fraud.

Column 2 shows results for the equation predicting fraudulent misrepresentations. These firms tend to have a history of regulatory violations, which can be interpreted in three ways. First, past regulatory violations may trigger greater scrutiny, which uncovers misrepresentations. Second, past wrongdoing indicates internal control problems or unethical behavior.

Third, firms must report past violations of all affiliated firms, and financial industry affiliations may create conflicts of interest which lead to misrepresentation. There is also a strong negative relation with Investment Company Act, which likely reflects the more stringent auditing and reporting requirements for mutual funds. Misrepresentation is also significantly lower for firms with a dedicated chief compliance officer, suggesting a benefit from internal monitoring.

Column 3 shows results for the equation predicting mutual fund late trading. We include mutual fund late trading as a separate category because it is quite different from the other types of fraud, and is unlikely to be repeated in the future. Mutual fund late trading is quite predictable, likely because it occurred only among a specific subsample - large mutual fund management firms. Perhaps the most important finding is not the predictability of mutual fund late trading itself, but that the predictability in the preceding columns shows that mutual fund late trading does not drive the results.

The last two columns show results for a multinomial probit model in which fraud is categorized as either firm-wide or committed by a rogue employee. Firm-wide frauds are committed by high level executives, or at the very least, with the implicit acceptance of the firm. Rogue employee fraud is committed by individuals who evade their firms' internal control systems and the firms do not knowingly benefit. While both types of fraud harm investors, fraud by a rogue employee is usually less costly and often the firm has sufficient assets to repay investors' losses. The results for firm-wide fraud are very similar to the results for all frauds reported in Table 4. At a false positive rate of 5% we are able to predict 24.2% of all firm-wide frauds, compared to 29.0% of all frauds. Thus, while firm-wide fraud is more difficult to predict, our results are not driven by rogue employees.

6 Are Investors Compensated for Fraud Risk?

The fact that certain characteristics predict fraud does not necessarily imply any problems with the current system of disclosure. Many characteristics that predict fraud may provide offsetting benefits for investors. For example, affiliation with a broker/dealer may reduce transaction costs or expedite trading. Lower internal monitoring may increase fraud risk but allow the firm to charge investors lower fees (e.g. Cassar and Gerakos (2010)). If higher fraud risk creates efficiency gains which are shared with investors through higher returns or lower fees, then the predictability of fraud does not imply any need for regulatory action provided there is clear disclosure and investors fully understand the tradeoffs.

Until this point in the paper, we have examined the entire sample of registered investment advisors. However, testing whether investors are compensated for fraud risk through performance or fees requires fund level data. As discussed in the data section, fund level data are available for only a minority of investment advisors. Thus the results in this section are less general and may not apply to firms for which we do not have return data.

On August 1^{st} of each year we measure each firm's fraud risk as the predicted value from the probit regression in the second column of Table 4, and assign this value to each fund managed by the firm. Using this measure of fraud risk, for each of the three fund level samples we divide funds into two equally weighted portfolios: funds managed by firms predicted to commit fraud within the next year, and the remainder.[9] (i.e. the high fraud risk portfolio contains fund managed by firms whose predicted fraud risk is above the 95^{th} percentile of clean firms). The remaining funds are placed in the second portfolio.

For the TASS hedge fund sample, we estimate alphas using the Fung and Hsieh (2001) eight factor model[10] over the 72 month period from August 2001 to July 2007. The results

[9] We have also divided firms into terciles based on the managing firm's fraud risk, with similar results

[10] We are grateful to David Hsieh for providing the factors used for these regressions. These factors are available on his website: `http://faculty.fuqua.duke.edu/~dah7/DataLibrary/TF-FAC.xls`.

in Panel A of Table 7 show that high fraud risk is associated with significantly *lower* risk adjusted returns.

For the CRSP mutual fund and PSN database samples we estimate alphas using the Carhart (1997) model over the 72 month period from August 2001 to July 2007. The results do not provide any evidence that fraud risk is compensated through higher alphas.

In Panel B of Table 7 we estimate the relation between the fees charged by a fund and an indicator variable for funds managed by firms whose fraud risk is above the 95^{th} percentile of the full sample of firms. Within each of the three fund samples, we estimate annual, pooled regressions with the standard errors clustered by firm and time. In all regressions we include indicator variables for the funds' style[11] and year fixed effects. We do not find any evidence that high fraud risk is associated with lower fees in any of the samples.

Taken together, the results in both panels suggest that investors are not compensated for fraud risk. Indeed, there is some evidence that fraud risk is associated with worse performance. This has important implications for interpreting the results in the previous section. If fraud risk were compensated, then investors could use our results to make an informed risk-reward tradeoff regarding fraud risk. Given that we do not find compensation, the implication is very clear; investors should avoid high fraud risk firms.

7 Implementation and Disclosure Format

In equilibrium, we would expect that fraud is either unpredictable or that investors are compensated for the predictable component. Thus, taken together, the findings in the previous two sections are puzzling, as they suggest investors voluntarily assume an uncompensated risk. Until now, however, we have ignored the cost and difficulty of actually estimating fraud risk. Although the SEC discloses each firm's current Form ADV filing, historical filings were

[11]We use the PrimaryCategory variable for the TASS hedge funds, Lipper categories for the CRSP mutual funds, and indicator variables for market capitalization and value/growth categories for the PSN funds.

not publicly available. In this section, we test whether the absence of historical Form ADV filings reduces investors' ability to predict fraud.

Table 8 shows the results of a fraud prediction model that uses only the publicly available cross-section of Form ADV filings at each point in time. For example, August 1^{st}, 2005 the independent variables are taken from each firm's last Form ADV filing prior to this date. We then estimate a backward looking cross-sectional probit regression, where the dependent variable equals one for all firms with a prior history of fraud (fraud occurring between January 1^{st}, 1996 and July 31^{st}, 2005 and discovered by the SEC before July 31^{st}, 2005). We then classify all firms in August 2005 into two groups based on the predicted values from this regression. Firms are classified as high risk if their predicted value is above the 95^{th} percentile of historically clean firms; the remainder are classified as low fraud risk. Panel B shows the percentage of frauds occurring between August 1^{st}, 2005 and July 31^{st}, 2006 that were committed by firms classified as high risk.

For comparison purposes, Table 9 shows the results of a model that predicts fraud using all prior Form ADV filings. For example, on August 1^{st}, 2005 the independent variables are taken from each firm's Form ADV filings as of August 1^{st}, 2001, 2002, 2003, and 2004. For each firm-year observation the dependent variable equals one if the firm commits a fraud during the subsequent 12 months. We then classify all firms on August 1^{st}, 2005 into two groups based on the predicted values from this regression. Firms are classified as high risk if their predicted value is above the 95^{th} percentile of the clean firm-year observations; the remainder are classified as low fraud risk. Panel B shows the percentage of frauds occurring between August 1^{st}, 2005and August 1^{st}, 2006 were committed by members of the high fraud risk group.

Our main interest in the regressions in Tables 8 and 9 is not the coefficient estimates, but rather in comparing the proportion of frauds correctly predicted. Across all years, using a 5% false positive rate to classify firms, the backward looking cross-sectional approach is still able to predict 25.9% of frauds compared to 31.4% for the panel approach. Although both models perform reasonably well, the panel regression is significantly better at predicting fraud out-of-sample (p-value<0.01). This result provides a partial explanation of why fraud risk is both predictable and uncompensated; during our sample period the SEC did not provide historical Form ADV data.

Another issue which has likely limited the usefulness of the information in Form ADV, is that historically Form ADV filings were not provided in a format amenable to statistical analysis. Investors had to manually download each Form ADV individually as an HTML encoded documents and extract the data. The cost of doing so may have outweighed the benefits for the vast majority of investors. Admati and Pfleiderer (2000) and Zingales (2009) argue that one of the key reasons for requiring disclosure is to provide a basis of comparison. After the first draft of this paper was circulated, the SEC began to provide historical Form ADV data in a standardized format (for 2008 onwards and some months in 2006 and 2007). This change should substantially increase investors' ability to use the information in Form ADV.

8 Conclusion

Overall, our results suggest that the current disclosures required by the SEC have significant power for predicting future fraud. If investors avoided the 5% of firms identified as having the highest ex ante predicted fraud risk in our sample, they could have avoided total dollar losses from fraud in excess of $4 billion. Based on the SEC's estimate of 9.01 hours to fill out Form ADV and an assumed cost of $1,000 per hour, during this same time period the direct costs of disclosure were at most $500 million. Thus, even ignoring the deterrent effect of Form ADV, this simple back-of-the-envelope calculation suggests that the benefits of Form ADV substantially outweigh the costs. Our results indicate that investors could avoid high fraud risk firms without sacrificing returns or paying higher fees. However, because the SEC did not make historical Form ADV data available to the investing public, the ability of investors to develop and use predictive models based on Form ADV data was potentially limited, and the realized benefits of disclosure during this time period may have been substantially lower.

Table 1

Summary of Investment Frauds

This table summarizes investment fraud committed between August 2001 and July 2010 as reported by the SEC's administrative actions and litigation releases. Registered denotes firms that file a Form ADV with the SEC. Firm-wide frauds are committed by high level executives, or at the very least, with the implicit acceptance of the firm. Rogue employee fraud is committed by individuals who evade their firms' internal control systems and the firm does not knowingly benefit. Ponzi schemes combine theft with payments of stolen money to early investors so as to evade detection and lure in new investors. Direct Theft also involves stealing money, but without the payments to early investors that characterizes a Ponzi scheme. Self Dealing includes all types of fraud in which an advisor trades securities on behalf of clients but illegally profits from their clients' trades. Overstating Assets occurs when an investment manager overstates returns or asset values, and charges unwarranted fees based on these inflated values. Mutual Fund Late Trading includes the well publicized cases in which mutual funds allowed certain investors to place trades after closing. Misrepresentation occurs when investment advisors lie to attract new investors (e.g. misrepresenting assets under management, or past regulatory violations).

Panel A: Registered vs. Non-Registered Advisers

	Total	Firm-Wide	Rogue Employee	Total Sum ($ billion)
Non-Registered	251	244	7	4.5
Registered	258	217	41	32.4
Total	509	461	48	36.9

Panel B: Classification of Fraud

		Amount ($ million)				Duration (years)		
Classification	Obs.	Mean	Median	Max	Missing	Mean	Median	Max
Ponzi	20	1,469.0	14.9	18,000.0	2	8.8	7.0	20.5
Direct Theft	74	43.4	2.6	554.0	5	5.5	4.7	20.9
Self Dealing	75	11.5	6.7	56.0	24	5.9	5.8	23.9
Overstating Assets	22	20.6	3.7	160.0	5	5.8	5.0	14.8
Mutual Fund Late Trading	29	56.2	28.5	250.0	0	5.4	4.7	11.6
Misrepresentation	38	44.1	2.1	415.0	28	5.5	5.5	11.1
Total	258	167.2	5.1	18,000.0	64	5.9	5.5	23.9

Table 2

Firm Summary Statistics

This table presents statistics of the 13,853 registered investment advisor firms that filed Form ADV from August 2001 through July 2006. Panel A summarizes firm characteristics. Employee Ownership is the aggregate employee ownership of the firm. Percent Client Agents is the percentage of clients that are agents for the owners of the assets. Panel B tabulates firm policies and characteristics. Past Regulatory equals one if the firm reports any past regulatory violations. Past Civil or Criminal equals one if the firm reports any past civil or criminal violations. Past Fraud equals one if the firm has previously been accused of fraud. Past Affiliated Fraud equals one if any of the firm's affiliates has previously been accused of fraud. Referral Fees equals one if the firm compensates any party for client referrals. Interest in Transactions equals one if the firm: recommends securities in which it has an ownership interest, serves as an underwriter, or has any other sales interest. Soft Dollars equals one if the firm receives benefits other than execution from a broker-dealer in connection with clients' trades. Broker in Firm equals one if the firm employs registered representatives of a broker-dealer. Custody equals one the firm has custody of clients' cash or securities. Investment Company Act equals one if the firm is registered under the Investment Company Act of 1940. Dedicated CCO equals one if the chief compliance officer has no other job title. Hedge Fund Client equals one if more than 75% of the firm's clients are hedge funds. The column Clean (Fraud) summarizes firm-years in which a fraud is not committed (is committed). The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels based on Fisher's exact test.

Panel A: Firm Characteristics

	Mean	SD	25th	Median	75th
Assets Under Management ($ million)	2,590	18,900	34	88	400
Average Account Size ($ thousand)	81,300	814,000	322	1,182	22,800
Firm Age (years)	8.2	8.8	1.4	5.1	12.9
Employee Ownership	67.1%	44.5	0.0	100.0	100.0
Percent Client Agents	26.9%	33.4	5.0	10.0	37.5

Panel B: Firm Policies and Characteristics

	Total	Clean	Fraud
Past Regulatory	13.4%	13.2	37.8***
Past Civil or Criminal	4.4%	4.3	15.2***
Past Fraud	0.5%	0.4	3.6***
Past Affiliated Fraud	2.9%	2.8	7.2***
Referral Fees	43.4%	43.2	62.6***
Interest in Transaction	32.2%	31.9	57.4***
Soft Dollars	57.8%	57.7	63.9***
Broker in Firm	40.3%	40.0	63.7***
Custody	26.0%	25.9	40.0***
Investment Company Act	12.1%	11.9	31.0***
Dedicated CCO	15.3%	15.3	17.9
Hedge Fund Client	10.5%	10.5	5.5***

Table 3

Fund Level Summary Statistics

This table presents summary statistics for the three subsamples of merged Form ADV - return database data. The first column presents characteristics of the merged Form ADV - TASS dataset (TASS). The second column presents characteristics of the merged Form ADV - CRSP mutual fund dataset (CRSPMF). The third column presents characteristics of the merged Form ADV - Plan Sponsor Network dataset (PSN). For the merged TASS dataset, we use all funds. For the CRSPMF and PSN datasets, we only include equity funds.

	TASS	CRSPMF	PSN
Firms Matched	37.5%	100	88.2
Funds Matched	27.0%	100	84.8
Fund Assets Matched	28.6%	100	89.9
Monthly Return	0.75%	0.85	0.89
Management Fee	1.43%	1.47	0.68
Incentive Fee	15.3%	N/A	N/A

Table 4

Predicting Fraud

Panel A shows the results of pooled probit regressions predicting fraud. The dependent variable equals one if the firm commits fraud in the subsequent year. In columns one and two, the full Form ADV sample is used. In column three, only firms with no prior frauds are included. In column four, only firms with no prior frauds and no frauds by their affiliates are included. In column five, only firms with no violations, no prior frauds and no frauds by their affiliates are included. In the interest of brevity we do not report coefficients for the constants. Standard errors are clustered by firm and year. Z-scores are reported in square brackets. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The columns in Panels B, C, and D correspond to the columns in Panel A. Panel B shows the proportion of frauds that could be predicted within sample. Panel C shows the out-of-sample performance of each model, using Form ADV filings in 2006 to predict frauds that occur during 2008 through 2010. Panel D shows the results from K-fold cross-validation tests.

Panel A: Predictors of Fraud					
	Full Sample	Full Sample	No Prior	No Affiliated	No Violations
Past Regulatory	0.284***	0.280***	0.280***	0.249***	
	[4.20]	[4.11]	[4.07]	[3.49]	
Past Civil or Criminal	0.191**	0.191**	0.210**	0.193*	
	[2.13]	[2.20]	[2.35]	[1.87]	
Past Fraud		0.327**			
		[2.07]			
Past Affiliated Fraud		-0.118	-0.116		
		[1.13]	[1.14]		
Referral Fees	0.100*	0.099*	0.098*	0.106**	0.139**
	[1.79]	[1.77]	[1.77]	[1.97]	[2.40]
Interest in Transaction	0.197***	0.198***	0.196***	0.209***	0.184**
	[2.89]	[2.90]	[2.85]	[3.05]	[2.24]
Soft Dollars	-0.051	-0.046	-0.041	-0.041	-0.073
	[0.89]	[0.81]	[0.73]	[0.71]	[1.10]
Broker in Firm	0.118**	0.120**	0.119**	0.117**	0.096
	[2.01]	[2.04]	[2.00]	[1.97]	[1.55]
Custody	0.094	0.094	0.083	0.089	0.028
	[1.43]	[1.45]	[1.26]	[1.38]	[0.33]
Investment Company Act	0.263***	0.266***	0.273***	0.278***	0.273***
	[3.29]	[3.31]	[3.47]	[3.46]	[2.83]
Dedicated CCO	-0.088	-0.087	-0.083	-0.102	-0.056
	[0.86]	[0.83]	[0.79]	[0.93]	[0.53]
Majority Employee Owned	0.009	0.004	0.012	0.011	0.033
	[0.11]	[0.05]	[0.15]	[0.13]	[0.37]
log (Avg. Acct Size)	-0.072***	-0.070***	-0.063***	-0.056***	-0.028
	[4.25]	[4.11]	[3.63]	[3.02]	[1.12]
Percent Client Agents	0.003***	0.003***	0.003***	0.003***	0.003***
	[3.91]	[3.88]	[3.74]	[3.48]	[2.91]
Hedge Fund Client	0.031	0.031	0.018	0.019	0.030
	[0.27]	[0.27]	[0.15]	[0.16]	[0.22]
log (AUM)	0.060***	0.058***	0.052***	0.046***	0.020
	[4.10]	[3.98]	[3.54]	[2.90]	[0.93]
log (Firm Age)	0.002	0.002	0.003	0.003	0.008
	[0.20]	[0.18]	[0.28]	[0.25]	[0.66]
Model Chi-Square	181.49***	198.31***	172.61***	157.60***	63.22***
Observations	53,994	53,994	53,739	52,228	45,920

Panel B: Within Sample Predictions					
	Full Sample	Full Sample	No Prior	No Affiliated	No Violations
# Fraud	517	517	498	464	310
Fraud Predicted	150	148	135	118	44
	29.0%	28.6	27.1	25.4	14.2
Fraud Not Predicted	367	369	363	346	266
	71.0%	71.4	72.9	74.6	85.8
# Clean Firms	53,447	53,447	53,739	51,764	45,610
Clean Firms Not Accused	50,804	50,804	50,579	49,176	43,330
	95.1%	95.1	94.1	95.0	95.0
Clean Firms Falsely Accused	2,673	2,673	2,662	2,588	2,280
	5.0%	5.0	5.0	5.0	5.0
Prop. of Total $ Losses Avoided	41.3%	42.5	33.3	30.0	7.9
Panel C: Out of Sample Predictions (2008-2010)					
# Fraud	27	27	25	24	18
Fraud Predicted	10	10	7	7	2
	37.0%	37.0	28.0	29.2	11.1
Fraud Not Predicted	17	17	18	17	16
	63.0%	63.0	72.0	70.8	88.9
# Clean Firms	10,356	10,356	10,291	10,002	8,912
Clean Firms Not Accused	9,838	9,838	9,778	9,505	8,467
	95.0%	95.0	95.0	95.0	95.0
Clean Firms Falsely Accused	518	518	513	497	445
	5.0%	5.0	5.0	5.0	5.0
Panel D: K-Fold Cross-Validation Out of Sample Predictions (2001-2007)					
Avg # Fraud Predicted	143.3	141.9	125.8	104.5	35.0
Avg % Fraud Predicted	27.7%	27.4	25.3	22.5	11.3
Stdev Fraud Predicted (#)	3.64	3.81	4.72	4.33	2.66
Min # Fraud Predicted	135	134	113	97	32
Max # Fraud Predicted	149	149	134	114	42
Avg # False Positives	2,669.2	2,669.2	2,583.3	2,657.5	2,275.8
Avg % False Positives	5.0%	5.0	5.0	5.0	5.0
Stdev False Positives	0.95	0.99	1.02	1.05	0.91
Min # False Positives	2,668	2,668	2,582	2,656	2,274
Max # False Positives	2,671	2,671	2,585	2,659	2,277

Table 5

Initiation vs. Continuation

Panel A shows the results of a multinomial probit regression predicting fraud. In the first column, the dependent variable equals one for firms that initiate a new fraud in the subsequent year. In the second column, the dependent variable equals one for firms that continue a preexisting fraud in the subsequent year. The excluded category is clean firms. The third column shows p-values from chi-square tests of the null hypothesis that the estimated coefficients are equal in both equations. In the interest of brevity we do not report coefficients for the constants. All significance tests are based on standard errors clustered by firm. Z-scores are reported in square brackets. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. Panel B shows the proportion of frauds that could be predicted within sample.

Panel A: Predicting Initiation versus Continuance of Fraud			
	Initiate	Continue	p-value Difference
Past Regulatory	0.682***	0.704***	
	[2.72]	[3.58]	0.932
Past Civil or Criminal	0.625**	0.381*	
	[2.06]	[1.68]	0.428
Referral Fees	0.899***	0.115	
	[3.46]	[0.68]	0.002
Interest in Transaction	0.447	0.521**	
	[1.64]	[2.47]	0.791
Soft Dollars	-0.347	-0.077	
	[1.44]	[0.41]	0.243
Broker in Firm	0.773***	0.268	
	[3.03]	[1.51]	0.037
Custody	0.000	0.288	
	[0.00]	[1.43]	0.297
Investment Company Act	0.542**	0.620***	
	[1.96]	[2.83]	0.793
Dedicated CCO	-0.329	-0.179	
	[1.15]	[0.97]	0.599
Majority Employee Owned	0.058	-0.031	
	[0.24]	[0.16]	0.741
log (Avg. Acct Size)	-0.218***	-0.171***	
	[3.83]	[4.01]	0.433
Percent Client Agents	0.008**	0.009***	
	[2.13]	[3.37]	0.962
Hedge Fund Client	0.252	-0.008	
	[0.46]	[0.02]	0.662
log (AUM)	0.199***	0.144***	
	[4.37]	[4.09]	0.250
log (Firm Age)	-0.062	0.031	
	[1.27]	[0.66]	0.120
Past Fraud	0.396	0.706*	
	[0.69]	[1.74]	0.607
Past Affiliated Fraud	-0.572	-0.200	
	[1.24]	[0.73]	0.413
Overall Model p-value Difference			0.115

Panel B: Frauds Predicted		
	Initiate	Continued
# Fraud	87	430
Fraud Predicted	33	112
	37.9%	26.0
Fraud Not Predicted	54	318
	62.1%	74.0
# Clean Firms	53,907	53,564
Clean Firms Not Accused	51,234	50,819
	95.0%	94.9
Clean Firms Falsely Accused	2,673	2,745
	5.0%	5.1

Table 6

Predicting Types of Fraud

The table reports estimates of multinomial probit models, where the dependent variable is listed above each column. In the first model, the dependent variables are Theft, Misrepresentation, Mutual Fund Late Trading, or no fraud (the omitted case). Theft includes Direct Theft, Ponzi, Self Dealing, and Overstate Assets. In the second model, the dependent variables are Firm-Wide, Rogue Employee, or no fraud (the omitted case). Panel A reports the estimates of these models. In the interest of brevity we do not report coefficients for the constants. Standard errors are clustered by firm and year. Z-scores are reported in square brackets. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. Panel B shows the proportion of frauds that could be predicted within sample. The sample contains 53,994 firm-year observations.

Panel A: Predicting Specific Types of Fraud					
	Theft	Misrepresent	Late Trading	Firm-Wide	Rogue
Past Regulatory	0.694***	1.169***	0.011	0.691***	0.879**
	[3.24]	[2.80]	[0.02]	[3.41]	[2.08]
Past Civil or Criminal	0.392	0.344	0.564	0.028	1.605***
	[1.61]	[0.56]	[0.90]	[0.11]	[3.24]
Referral Fees	0.291*	0.019	0.144	0.193	0.933
	[1.65]	[0.04]	[0.20]	[1.14]	[1.61]
Interest in Transaction	0.452*	0.288	1.952	0.435**	1.183*
	[1.94]	[0.71]	[1.60]	[2.11]	[1.67]
Soft Dollars	-0.302	0.272	1.277	-0.029	-0.943*
	[1.52]	[0.61]	[1.61]	[0.15]	[1.92]
Broker in Firm	0.216	0.542	1.879*	0.326*	0.873
	[1.12]	[1.24]	[1.91]	[1.84]	[0.93]
Custody	0.300	0.108	-0.115	0.335*	-0.734
	[1.35]	[0.32]	[0.20]	[1.71]	[1.50]
Investment Company Act	0.767***	-1.118**	1.028**	0.650***	0.349
	[3.26]	[2.29]	[1.97]	[3.01]	[0.73]
Dedicated CCO	-0.144	-1.448***	0.376	-0.32	0.295
	[0.72]	[3.40]	[0.71]	[1.63]	[0.66]
Majority Emp. Owned	0.072	0.203	-1.754*	-0.007	0.04
	[0.34]	[0.46]	[1.72]	[0.04]	[0.07]
log (Avg. Acct Size)	-0.198***	-0.069	-0.120	-0.143***	-0.298***
	[4.73]	[0.50]	[1.24]	[3.22]	[3.32]
Percent Client Agents	0.009***	0.003	0.006	0.009***	0.003
	[3.35]	[0.54]	[0.88]	[3.43]	[0.45]
Hedge Fund Client	0.139	-0.184	-13.020***	0.002	-0.438
	[0.33]	[0.22]	[19.63]	[0.01]	[0.40]
log (AUM)	0.155***	0.199	0.148	0.119***	0.369**
	[4.43]	[1.53]	[1.53]	[3.25]	[2.48]
log (Firm Age)	-0.007	0.083	0.087	0.004	0.186
	[0.14]	[0.85]	[0.41]	[0.08]	[0.95]
Past Fraud	0.618	0.755	0.802	-0.068	1.033*
	[1.29]	[0.83]	[0.70]	[0.11]	[1.87]
Past Affiliated Fraud	-0.076	-0.236	-1.725**	-0.483	0.234
	[0.27]	[0.36]	[2.49]	[1.50]	[0.46]
Overall Chi-Square		4580.38***		344.37***	

Panel B: Within Sample Predictions					
	Theft	Misrepresent	Late Trading	Firm-Wide	Rogue
# Fraud	390	83	44	450	67
Fraud Predicted	107	19	21	109	50
	27.4%	22.9	47.7	24.2	74.6
Fraud Not Predicted	283	64	23	341	17
	72.6%	77.1	52.3	75.8	25.4
Clean Firms	53,477	53,477	53,477	53,477	53,477
Clean Firms Not Accused	50,804	50,804	50,804	50,804	50,804
	95.0%	95.0	95.0	95.0	95.0
Clean Firms Falsely Accused	2,673	2,673	2,673	2,673	2,673
	5.0%	5.0	5.0	5.0	5.0

Table 7

Fraud Risk, Alphas, and Fees

In this table we test the relation of fraud risk with alphas and fees. We merge our Form ADV sample with the TASS hedge fund database (TASS), CRSP Mutual Fund database (CRSPMF), and PSN institutional fund database (PSN). Fraud risk is the predicted values from the regressions reported in column two of Table 4. We rank all firms in the full Form ADV sample by fraud risk. In Panel A, for each database we form two portfolios: low fraud risk for funds from firms whose fraud risk is below the 95^{th} percentile of all firms and high fraud risk for funds from firms whose fraud risk is above the 95^{th} percentile of all firms. Portfolios are equal weighted. We estimate alphas using monthly returns for each portfolio. We use the Fung and Hsieh (1997) model for the TASS sample and the Carhart (1997) model for CRSPMF and PSN samples. High-Low is the alpha of a portfolio long high fraud risk funds and short low fraud risk funds. The t-statistics, reported in square brackets, are adjusted using the method of Newey and West (1987) with three lags. Panel B reports the relation between fraud risk and fees. The dependent variables for the TASS sample are the reported management and incentive fees. The dependent variable for the CRSP mutual fund sample is expense ratios. The dependent variable for the PSN sample is the reported fee percentage charged on a $50 million account. High risk equals one if the firm-level predicted value from column two of Table 4 is at or above the 95^{th} percentile of all firms. The variables are measured as of August 1^{st} of each year. The standard errors are clustered by firm. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Portfolio Alphas

	Low Fraud Risk	High Fraud Risk	High-Low
TASS	0.005***	0.003***	-0.001**
	[5.92]	[3.86]	[2.06]
CRSP	-0.206***	-0.211***	-0.005
	[5.46]	[4.49]	[0.29]
PSN	0.007	0.013	0.006
	[0.13]	[0.19]	[0.21]

Panel B: Fees				
	TASS		CRSPMF	PSN
	Management Fee	Incentive Fee	Expense Ratio	Management Fee
High Risk	-0.08	0.19	0.00	-0.01
	[1.23]	[0.27]	[0.60]	[0.81]
Style Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	7,647	7,647	23,211	12,442
R^2	0.10	0.36	0.01	0.17

Table 8
Point-In-Time Tests Using Publicly Available Data

Panel A shows the estimates from backward looking cross-sectional probit regressions. The dependent variable equals one for firms which have a prior history of fraud (fraud occurring between January 1996 and August 1^{st} of the year in which the independent variables are observed). The independent variables reflect the publicly available data as of August 1^{st} of each year. In the interest of brevity we do not report coefficients for the constants. Standard errors are robust. Z-scores are reported in square brackets. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The columns in Panel B correspond to the columns in Panel A. In Panel B, the model is used to predict frauds occurring in the next year.

Panel A: Backward Looking Cross-Sections						
	Aug '01	Aug '02	Aug '03	Aug '04	Aug '05	Aug '06
Past Regulatory	0.717***	0.990***	0.994***	0.836***	0.952***	1.171***
	[4.17]	[4.85]	[6.78]	[5.45]	[7.56]	[8.78]
Past Civil or Criminal	0.657***	0.457**	0.558***	0.607***	0.391**	0.458***
	[3.07]	[2.41]	[3.23]	[3.61]	[2.49]	[3.49]
Referral Fees	0.442**	0.355**	0.323**	0.223	0.13	0.125
	[2.25]	[1.98]	[2.07]	[1.57]	[1.11]	[1.06]
Interest in Transaction	0.346	0.434**	0.022	-0.236	0.061	0.081
	[1.51]	[2.06]	[0.12]	[1.37]	[0.43]	[0.65]
Soft Dollars	-0.333*	-0.250	-0.197	-0.332**	-0.084	0.026
	[1.90]	[1.61]	[1.25]	[2.22]	[0.70]	[0.21]
Broker in Firm	0.134	0.067	-0.085	-0.203*	-0.154	-0.132
	[0.78]	[0.37]	[0.67]	[1.69]	[1.26]	[1.05]
Custody	-0.016	0.045	-0.001	0.183	-0.077	0.009
	[0.09]	[0.26]	[0.01]	[1.27]	[0.64]	[0.08]
Investment Company Act	0.153	0.095	0.014	0.370**	0.157	0.169
	[0.68]	[0.44]	[0.07]	[2.21]	[0.88]	[1.04]
Dedicated CCO	-0.385	-0.21	-0.082	0.179	0.072	0.117
	[1.03]	[0.90]	[0.34]	[1.11]	[0.64]	[1.11]
Majority Emp. Owned	-0.127	0.072	-0.004	-0.065	0.085	0.144
	[0.58]	[0.40]	[0.02]	[0.45]	[0.70]	[1.14]
log (Avg. Acct Size)	-0.151***	-0.108***	-0.056	-0.073**	-0.094***	-0.077**
	[3.14]	[3.01]	[1.40]	[2.13]	[3.67]	[2.45]
Percent Client Agents	0.001	-0.001	-0.001	0.002	0.002	0.0001
	[0.45]	[0.40]	[0.44]	[0.67]	[0.70]	[0.05]
Hedge Fund Client	0.240	-0.060	-0.090	-0.116		-0.413
	[0.54]	[0.15]	[0.23]	[0.30]		[1.06]
log (AUM)	0.116***	0.095***	0.049	0.068**	0.076***	0.057**
	[3.13]	[2.90]	[1.57]	[2.29]	[3.24]	[2.19]
log (Firm Age)	0.050*	0.050	0.158***	0.147***	0.127**	0.191***
	[1.66]	[1.12]	[2.94]	[2.70]	[2.32]	[3.88]
Past Affiliated Fraud	0.484*	0.583**	0.034	-0.061	-0.053	-0.177
	[1.78]	[2.49]	[0.12]	[0.26]	[0.26]	[0.86]
Model Chi-Square	245.12***	124.92***	120.93***	157.89***	151.70***	160.32***
Observations	7,352	7,747	8,562	9,088	10,862	10,383

Panel B: Out-of-Sample Predictions						
	Aug '01	Aug '02	Aug '03	Aug '04	Aug '05	Aug '06
# Fraud	104	116	115	83	59	40
Fraud Predicted	31	35	29	20	11	10
	29.8%	30.2	25.2	24.1	18.6	25.0
Fraud Not Predicted	73	81	86	63	48	30
	70.2%	69.8	74.8	75.9	81.4	75.0
# Clean Firms	7,248	7,631	8,447	9,005	10,803	10,343
Clean Firms Not Accused	6,890	7,258	8,030	8,549	10,239	9,791
	95.1%	95.1	95.1	94.9	94.8	94.7
Clean Firms Falsely Accused	358	373	417	456	564	552
	4.9%	4.9	4.9	5.1	5.2	5.3

Table 9

Predictions Using a Panel of All Prior Years

Panel A shows the results of fraud prediction models that use all prior Form ADV filings to predict fraud For each firm-year observation the dependent variable equals one if the firm commits a fraud during the subsequent 12 months. Years are as of August 1^{st} of each year. In the interest of brevity we do not report coefficients for the constants. Standard errors are clustered by firm and year. Z-scores are reported in square brackets. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. Panel B shows the proportion of frauds in year N+1 that could be predicted using all data in years up to N-1 and then forming a prediction model based on frauds in year N.

Panel A: Panel of All Prior Years						
	Aug '01	Aug '02	Aug '03	Aug '04	Aug '05	Aug '06
Past Regulatory	0.186*	0.211***	0.259***	0.287***	0.291***	0.280***
	[1.88]	[3.51]	[3.40]	[3.80]	[4.13]	[4.11]
Past Civil or Criminal	0.237*	0.219***	0.171*	0.203**	0.174*	0.191**
	[1.73]	[2.59]	[1.72]	[2.16]	[1.88]	[2.20]
Referral Fees	0.040	0.029	0.043	0.060	0.083	0.099*
	[0.44]	[0.63]	[0.81]	[1.13]	[1.51]	[1.77]
Interest in Transaction	0.266***	0.257***	0.249***	0.227***	0.217***	0.198***
	[2.79]	[4.90]	[4.27]	[3.43]	[3.29]	[2.90]
Soft Dollars	-0.035	-0.051	-0.040	-0.033	-0.033	-0.046
	[0.38]	[1.01]	[0.70]	[0.60]	[0.61]	[0.81]
Broker in Firm	0.201**	0.161***	0.127**	0.117*	0.111*	0.120**
	[2.31]	[2.87]	[2.05]	[1.93]	[1.88]	[2.04]
Custody	0.001	0.043	0.090	0.072	0.086	0.094
	[0.01]	[0.64]	[1.14]	[1.02]	[1.32]	[1.45]
Investment Company Act	0.242**	0.287***	0.300***	0.302***	0.274***	0.266***
	[2.39]	[4.08]	[4.55]	[4.64]	[3.56]	[3.31]
Dedicated CCO	0.252	0.308***	0.348***	0.198	0.025	-0.087
	[1.56]	[3.16]	[3.49]	[1.26]	[0.20]	[0.83]
Majority Emp. Owned	-0.085	-0.097*	-0.051	-0.003	0.017	0.004
	[0.85]	[1.86]	[0.69]	[0.03]	[0.21]	[0.05]
log (Avg. Acct Size)	-0.098***	-0.091***	-0.085***	-0.078***	-0.071***	-0.070***
	[3.80]	[7.71]	[5.81]	[4.69]	[4.02]	[4.11]
Percent Client Agents	0.004***	0.004***	0.004***	0.003***	0.003***	0.003***
	[3.01]	[4.17]	[4.26]	[3.89]	[3.94]	[3.88]
Hedge Fund Client	0.009	0.072	0.112	0.107	0.053	0.031
	[0.03]	[0.53]	[0.86]	[0.83]	[0.42]	[0.27]
log (AUM)	0.089***	0.083***	0.074***	0.066***	0.060***	0.058***
	[4.16]	[7.67]	[5.44]	[4.41]	[3.89]	[3.98]
log (Firm Age)	0.018	0.014	0.008	0.008	0.007	0.002
	[0.97]	[1.41]	[0.65]	[0.69]	[0.68]	[0.18]
Past Fraud	0.402	0.371*	0.285	0.289	0.327*	0.327**
	[1.32]	[1.75]	[1.20]	[1.51]	[1.88]	[2.07]
Past Affiliated Fraud	-0.206	-0.208*	-0.193*	-0.157	-0.134	-0.118
	[0.99]	[1.67]	[1.87]	[1.56]	[1.24]	[1.13]
Model Chi-Square	139.16***	545.20***	307.33***	218.63***	205.51***	198.31***
Observations	7,352	15,099	23,661	32,749	43,611	53,994

Panel B: Out of Sample Predictions (Year T+2)						
	Aug '01	Aug '02	Aug '03	Aug '04	Aug '05	Aug '06
# Fraud	116	115	83	59	40	27
Fraud Predicted	47	38	21	12	11	9
	40.5%	33.0	25.3	20.3	27.5	33.3
Fraud Not Predicted	69	77	62	47	29	18
	59.5%	67.0	74.7	79.7	72.5	66.7
# Clean Firms	7,631	8,447	9,005	10,803	10,343	10,356
Clean Firms Not Accused	7,231	8,049	8,495	10,174	9,818	9,893
	94.8%	95.3	94.3	94.2	94.9	95.5
Clean Firms Falsely Accused	400	398	510	629	525	463
	5.2%	4.7	5.7	5.8	5.1	4.5

Figure 1
One Fraud Case's Timeline

This figure shows the timeline of one particular fraud, committed by K.W. Brown & Company, from initiation to the end of all legal actions. Beginning in September 2002 the firm began defrauding clients through self-dealing. The firm traded securities for its own proprietary account as well as on behalf of clients. The firm engaged in ex post allocation of trades; securities were purchased but not allocated to a specific account. At a later date, profitable trades were retroactively allocated to the firm's proprietary account and unprofitable trades were allocated to clients. This resulted in over $4.5 million in illegal gains for the firm, and more than $9 million in client losses.

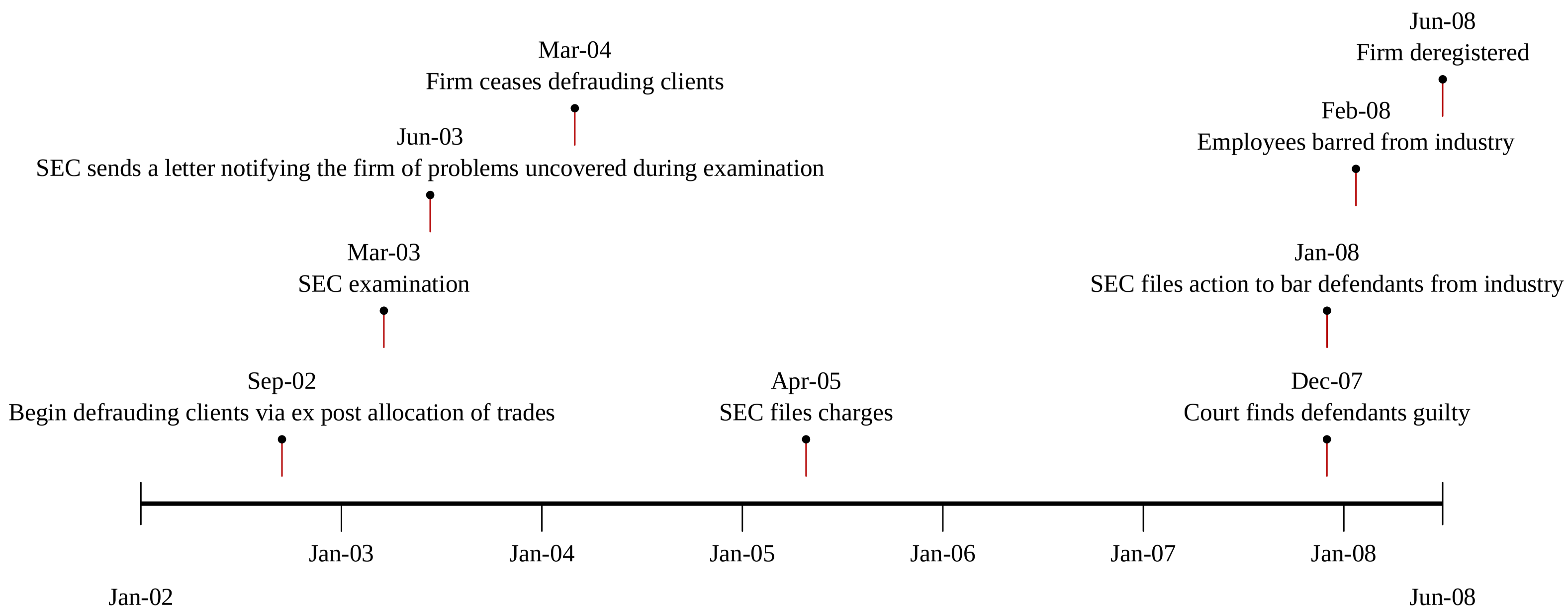
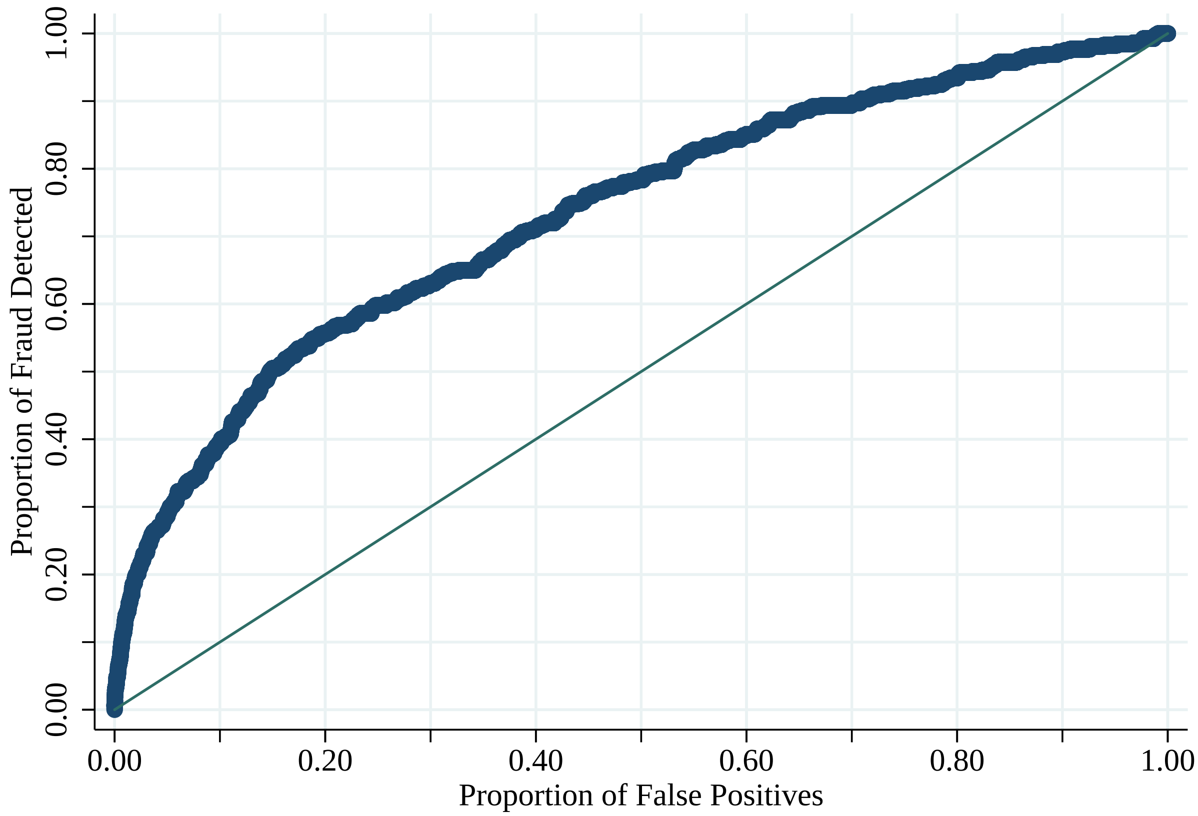


Figure 2

Proportion of Fraud Predicted for All False Positive Rates

This figure shows the receiver operating characteristic (ROC) curve for the probit regression results from the second column of Table 4. The ROC curve shows the relation between the proportion of frauds detected and the proportion of false positives for all possible classification cutpoints. The ROC curve is generated by taking each observation's estimated fraud probability, computing the sensitivity and false positives using that point as a cutoff, and then plotting the results.



Appendix 1
Variable Definitions

This table contains variable definitions and details of their construction from the Form ADV data.

Variable	Definition	Form ADV Variables
Past Regulatory	Filed a regulatory DRP	One or more of the following items: 11c1, 11c2, 11c3, 11d1, 11d2, 11d3, 11d4, 11d5, 11e1, 11e2, 11e3, 11e4
Past Civil or Criminal	Filed a criminal or civil DRP	One or more of the following items: 11a1, 11a2, 11b1, 11b2, 11h1a, 11h1b, 11h1c, 11h2
Past Fraud	The firm committed a publicly observed fraud	SEC Administrative Proceeding or Litigation Release was filed for firm prior to August 1^{st} of firm-year observation.
Past Affiliated Fraud	An affiliate of the firm listed on Schedule D Section 7.A committed a publicly observed fraud	SEC Administrative Proceeding or Litigation Release was filed for affiliated firm prior to August 1^{st} of firm-year observation & Form ADV Schedule D Section 7.A reports fraud firm as affiliate.
Referral Fees	Do you or any related person, directly or indirectly, compensate any person for client referrals?	Item 8f
Interest in Transaction	Do you or any related person: (1) buy securities for yourself from advisory clients, or sell securities you own to advisory clients (2) recommend securities (other than investment products) to advisory clients in which you or any related person has some other proprietary (ownership) interest (3) recommend purchase of securities to advisory clients for which you or any related person serves as underwriter, general or managing partner, or purchaser representative (4) recommend purchase or sale of securities to advisory clients for which you or any related person has any other sales interest	One or more of the following items: 8a1, 8a3, 8b2, 8b3

Appendix 1: Variable Definitions (continued)

Soft Dollars	Do you or any related person receive research or other products or services other than execution from a broker-dealer or a third party in connection with client securities transactions?	Item 8e
Broker in Firm	Number of registered representatives of broker-dealers employed by firm is greater than zero	5b2>0
Custody	Do you or any related person have custody of any advisory clients' cash or securities?	One or more of the following items: 9a1, 9a2, 9b1, 9b2
Investment Company Act	You are an investment adviser (or sub-adviser) to an investment company registered under the Investment Company Act of 1940	Item 2a4
Dedicated CCO	CCO has no other stated role within firm	Officer listed as CCO on Schedule A has no other "Title or Status"
Majority Employee Owned	Over 50% aggregate employee ownership	See Dimmock, Gerken, and Marietta-Westberg (2010) for imputation method
log (Avg. Acct Size)	Logarithm of reported assets under management divided by number of investors (plus one)	log (Item 5f2c/(Item 5f2f+1)+1)
Percent Client Agents	Sum of percent of banking/thrift, mutual, pension, charitable, corporate, and government clients	Imputation method sames as Dimmock, Gerken, and Marietta-Westberg (2010) then add Items: 5d3, 5d4, 5d5, 5d7, 5d8, and 5d9
Hedge Fund Client	Primarily hedge fund clients	Item 5d6 $\geq$ 75%
log (AUM)	Logarithm of assets under management	log (Item 5f2c+1)
log (Firm Age)	Logarithm of firm age in years	log (number of days since date registration with the SEC became effective)

Appendix 2
Length of Fraud

This table presents Tobit regression estimates where the dependent variable is the logarithm of the length of the fraud in years. The full sample includes one observation per fraud with sufficient data to calculate duration.

	Full Sample	Full Sample	No Prior	No Affiliated	No Violations
Past Regulatory	0.026	0.048	0.014	0.055	
	[0.19]	[0.35]	[0.11]	[0.39]	
Past Civil or Criminal	0.065	0.064	0.089	0.073	
	[0.38]	[0.37]	[0.48]	[0.39]	
Past Fraud		0.101			
		[0.36]			
Past Affiliated Fraud		-0.206		-0.286	
		[0.98]		[1.28]	
Referral Fees	-0.155	-0.150	-0.166	-0.156	-0.179
	[1.24]	[1.20]	[1.30]	[1.23]	[1.08]
Interest in Transaction	-0.087	-0.079	-0.069	-0.059	-0.155
	[0.74]	[0.67]	[0.57]	[0.49]	[1.03]
Soft Dollars	0.064	0.067	0.057	0.061	0.086
	[0.55]	[0.57]	[0.47]	[0.50]	[0.56]
Broker in Firm	-0.112	-0.113	-0.119	-0.117	-0.108
	[0.90]	[0.90]	[0.94]	[0.93]	[0.72]
Custody	0.147	0.149	0.132	0.147	0.231
	[1.26]	[1.27]	[1.11]	[1.23]	[1.42]
Investment Company Act	0.036	0.043	0.004	0.018	0.008
	[0.23]	[0.27]	[0.03]	[0.11]	[0.04]
Dedicated CCO	0.060	0.074	0.081	0.098	0.061
	[0.47]	[0.57]	[0.60]	[0.73]	[0.33]
Majority Employee Owned	0.171	0.168	0.173	0.165	0.081
	[1.30]	[1.28]	[1.30]	[1.24]	[0.49]
log (Avg. Acct Size)	0.007	0.007	0.009	0.005	-0.033
	[0.22]	[0.23]	[0.28]	[0.17]	[0.63]
Percent Client Agents	-0.001	-0.001	-0.001	-0.001	-0.001
	[0.45]	[0.40]	[0.41]	[0.30]	[0.29]
Hedge Fund Client	-0.289	-0.291	-0.293	-0.285	-0.244
	[1.14]	[1.15]	[1.14]	[1.11]	[0.76]
log (AUM)	-0.002	-0.002	-0.004	0.000	0.028
	[0.09]	[0.09]	[0.15]	[0.02]	[0.65]
log (Firm Age)	0.070**	0.070**	0.071**	0.071**	0.095**
	[2.26]	[2.27]	[2.25]	[2.26]	[2.49]
Observations	182	182	175	175	116

References

Admati, Anat R., and Paul Pfleiderer, 2000, Forcing Firms to Talk: Financial Disclosure Regulation and Externalities, *Review of Financial Studies* 13, 479–519.

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

Bollen, Nicolas P.B., and Veronika K. Pool, 2008, Conditional Return Smoothing in the Hedge Fund Industry, *Journal of Financial and Quantitative Analysis* 43, 267–298.

Bollen, Nicolas P.B., and Veronika K. Pool, 2009, Do Hedge Fund Managers Misreport Returns? Evidence from the Pooled Distribution., *Journal of Finance* 64, 2257–2288.

Bollen, Nicolas P.B., and Veronika K. Pool, 2010, Predicting Hedge Fund Fraud with Performance Flags, Working Paper, Vanderbilt University.

Brown, Stephen J., William N. Goetzmann, Bing Liang, and Christopher Schwarz, 2008, Mandatory Disclosure and Operational Risk: Evidence from Hedge Fund Registration, *Journal of Finance* 63, 2785–2815.

Brown, Stephen J., William N. Goetzmann, Bing Liang, and Christopher Schwarz, 2009, Estimating Operational Risk for Hedge Funds: The ω-Score, *Financial Analysts Journal* 65, 43–53.

Brown, Stephen J., William N. Goetzmann, Bing Liang, and Christopher Schwarz, 2010, Trust and Delegation, Working Paper, New York University.

Busse, Jeffrey A., Amit Goyal, and Sunil Wahal, 2010, Performance Persistence in Institutional Investment Management, *Journal of Finance* 65, 765–790.

Carhart, Mark M., 1997, On Persistence in Mutual Fund Performance, *Journal of Finance* 52, 57–82.

Cassar, Gavin, and Joesph Gerakos, 2010, Determinants of Hedge Fund Internal Controls and Fees, *Accounting Review* forthcoming.

Dechow, Patricia M., Weili Ge, Chad R. Larson, and Richard G Sloan, 2010, Predicting Material Accounting Misstatements, *Contemporary Accounting Research* forthcoming.

Dimmock, Stephen G., William C. Gerken, and Jennifer Marietta-Westberg, 2010, Owners as Stewards: Employee Ownership of Institutional Investment Management Firms, Working Paper, Nanyang Technological University.

Fung, William, and David A. Hsieh, 2001, The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers, *Review of Financial Studies* 14, 313–341.

Greene, William H., 2003, *Econometric Analysis 5th edn.* (Prentice Hall, New Jersey).

Karpoff, Jonathan M., and John R. Lott, 1993, The Reputational Penalty Firms Bear from Committing Criminal Fraud, *Journal of Law and Economics* 36, 757–802.

Newey, Whitney K., and Kenneth D. West, 1987, A Simple, Positive Semi-Definite, Hetroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* 55, 703–708.

Straumann, Daniel, 2008, Measuring the quality of hedge fund data, *RiskMetrics Journal* Winter, 65–93.

Zingales, Luigi, 2009, The Future of Securities Regulation, *Journal of Accounting Research* 47, 391–425.

Zitzewitz, Eric, 2006, How Widespread Was Late Trading in Mutual Funds?, *American Economic Review* 96, 284–289.