Title	Predicting fraud by investment managers
Author(s)	Dimmock, Stephen G.; Gerken, William Christopher
Citation	Dimmock, S. G., & Gerken, W. C. (2012). Predicting Fraud by Investment Managers. Journal of Financial Economics, 105(1), 153-173.
Date	2012
URL	http://hdl.handle.net/10220/17824
Rights	© 2012 Elsevier B.V. This is the author created version of a work that has been peer reviewed and accepted for publication by Journal of Financial Economics, Elsevier B.V. It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [http://dx.doi.org/10.1016/j.jfineco.2012.01.002].

Predicting fraud by investment managers[☆]

Stephen G. Dimmock[a,*], William C. Gerken[b]

[a] *Division of Finance and Banking, Nanyang Technological University, Singapore, 639798*
[b] *Department of Finance, Auburn University, Auburn, AL 36849*

Abstract

We test the predictability of investment fraud using a panel of mandatory disclosures filed with the SEC. We find that disclosures related to 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 would allow an investor to avoid 29% of fraud cases and over 40% of the total dollar losses from fraud. We find no evidence that investors receive compensation for fraud risk through superior performance or lower fees. We examine the barriers to implementing fraud prediction models and suggest changes to the SEC's data access policies that could benefit investors.

Keywords: Fraud, Investment fraud, Operational risk, SEC, Disclosure, Form ADV
JEL: G2, G20, G28, K2, K22

[☆]We are grateful for comments from: two anonymous referees, James Angel, Devraj Basu, Ranadeb Chaudhuri, Elroy Dimson, Will Goetzmann, Philip Hamill, Chuan-Yang Hwang, Dusan Isakov, Zoran Ivkovic, Jun-Koo Kang, Ayla Kayhan, Naveen Khanna, Clive Lennox, Antonio Macias, Dinah McNichols, Jay Patel, Josh Pollet, Melvyn Teo, Fan Yu, Lei Zhang, seminar participants at West Virginia University, Singapore Monetary Authority, 2010 Asian FMA, 2010 CRSP Forum, Current Topics in Securities Regulation Conference, Emerging Scholars in Banking and Finance, 2010 EFA, 2010 FMA, Fourth Singapore International Conference, One-Day Conference on Professional Asset Management, and Oklahoma Risk Management conferences. We are especially grateful to Scott Weisbenner for helpful discussions. We acknowledge the financial support of Networks Financial Institute and its award for best paper in Financial Services Regulatory Reform.

[*]Corresponding Author. Tel: 65-6790-6119 Fax: 65-6791-3697

Email addresses: `dimmock@ntu.edu.sg` (Stephen G. Dimmock), `will.gerken@auburn.edu` (William C. Gerken)

August 7, 2011

1. Introduction

On December 11, 2008 the Securities and Exchange Commission (SEC) charged Bernard Madoff with securities fraud for committing an $18 billion Ponzi scheme.[1] This case emphasized the opportunities advisers have to exploit investors and the importance of limiting advisers' opportunistic behavior through either market or regulatory forces. In the U.S., the regulatory system protects investors primarily through mandatory disclosures. Investment advisers must file Form ADV to disclose information about their operations, conflicts of interest, disciplinary histories, and other material facts. Investors are then responsible for using these disclosures to assess advisers' fraud risk. In this paper, we address the question: Could investors use these mandatory disclosures to predict fraud?

To address this question, we use an annual panel of Form ADVs filed from August 2001 through July 2006. The panel includes 13,853 investment advisers who advise more than 20 million clients and control more than $32 trillion in assets (as of August 2005). These firms advise all mutual funds, nearly all institutional investment funds, and many hedge funds in the U.S. Although the SEC provides public access to each investment adviser's current Form ADV filing, this panel of historical filings is not publicly available, and we are the first researchers to use these data. Our data also include a review of all SEC administrative proceedings and litigation releases from August 2001 through July 2010 to identify those cases in which investment advisers defrauded their clients.

We find that Form ADV disclosures related to past regulatory violations, conflicts of interest, and monitoring are all significant predictors of fraud. Of key importance for investors and regulators, the results show that an investor who avoided the 5% of firms with the highest ex ante predicted fraud risk would avoid 29% of fraud cases and over 40% of the dollar losses from fraud[2] (although to obtain these benefits the investor would have to forgo investing with 5% of non-fraudulent advisers). Out-of-sample tests confirm the robustness of the fraud predictions.

These findings are subject to several limitations. First, only *detected* fraud

---

[1] See "SEC Charges Bernard L. Madoff for Multi-Billion Dollar Ponzi Scheme": `http://www.sec.gov/news/press/2008/2008-293.htm`.

[2] For example, the predicted fraud risk of Bernard L. Madoff Investment Securities is above the 95^{th} percentile.

cases are included in the prediction models. Although we conduct extensive out-of-sample tests, we cannot reject the possibility that prediction models are biased because undetected fraud cases are unobservable. Second, although we find that certain characteristics, such as conflicts of interest, can predict fraud, we cannot infer that conflicts of interest cause fraud, or that their prohibition would deter fraud. Prediction does not imply causality, as firms' characteristics may be jointly determined with the decision to commit fraud. Third, in addition to the disclosures mandated by the SEC, investors may assess fraud risk using other sources of information that we do not include in our models. Finally, prediction is not the sole purpose of disclosure; it is also intended to deter fraud. We do not address this deterrent effect of disclosure in this paper.

If the Form ADV data were not useful for predicting fraud, then either disclosure deters fraud so effectively that it eliminates the predictability that would occur in the absence of disclosure or the disclosed information is worthless. Our findings thus provide evidence that regulators require investment advisers to disclose relevant information.

The predictability of fraud raises the question: why do investors allocate money to firms with high fraud risk? One possibility is that the characteristics that predict fraud provide offsetting benefits for investors. For example, affiliation with a brokerage firm could reduce transaction costs or expedite trading. In-house custody of clients' assets could increase fraud risk but reduce costs, resulting in lower fees for investors [e.g., Cassar and Gerakos (2010)]. Darby and Karni (1973), Karpoff and Lott (1993), Klein and Leffler (1981), and Lott (1996) argue that if investors differ in their valuation of fraud risk then some investors would accept a high level of fraud risk in return for superior performance or lower fees, while other investors would choose low fraud risk and accept worse performance or higher fees. To test whether investors receive compensation for fraud risk, we classify investment funds based on their advisers' predicted fraud risk. This subsample includes only the subset of firms that manage funds included in the TASS hedge fund, CRSP mutual fund, and/or PSN Informa databases. For all three types of funds, we find no evidence that investors receive compensation for fraud risk through superior performance or lower fees. However, we cannot rule out the possibility that investors receive some other form of compensation.

Given the surprising result that fraud risk is both predictable and apparently uncompensated, we turn to another possibility. Perhaps barriers to implementing a predictive model cause the costs to outweigh the potential benefits. To explore this possibility, we compare two types of predictive models, both of which take the perspective of an investor attempting to implement a fraud prediction model during the sample period. The first predicts fraud using only the limited subset of information that would have been publicly accessible. Until 2010, the general public had access to only contemporaneous cross-sections of filings; thus, the independent variables in these tests are taken from the contemporaneously accessible filings. The second type of model predicts fraud using a panel of prior filings. These tests show what would have been possible if historical filings have been contemporaneously accessible during the sample period; these models are moderately better at predicting fraud out-of-sample. We discuss simple changes to data access policies that could improve investors' ability to predict fraud.

The paper proceeds as follows. Section 2 discusses the related literature. Section 3 describes our data. Section 4 contains tests of the predictability of fraud. Section 5 tests the relation of fraud risk with the performance and fees of investment funds. Section 6 examines the costs and barriers to predicting fraud. Section 7 concludes.

2. Related research

To our knowledge, just two papers, Bollen and Pool (2010) and Zitzewitz (2006), develop methods to detect fraud by investment advisers. Bollen and Pool (2010) build on earlier studies of hedge funds' manipulation of reported returns [Bollen and Pool (2008); Bollen and Pool (2009); and Straumann (2009)] and find that suspicious return patterns can predict fraud charges. Zitzewitz (2006) shows that daily fund flows provide information about late trading in mutual funds. Although these papers, like ours, develop methods to detect fraud, they analyze returns and fund flows rather than firms' disclosures of business practices and conflicts of interest. An advantage of using firms' disclosures is that we can actually predict fraud, whereas methods based on returns and flows can only detect past or ongoing fraud. A further advantage is that Form ADV disclosures are mandatory, whereas the disclosure of returns is optional for many investment advisers.

Brown et al. (2008, 2009) examine operational risk using a cross-section of Form ADV filings from hedge fund advisers. The authors define "problem" funds as those managed by an adviser that reports prior legal or regulatory violations, that are either committed by the adviser itself or an affiliated firm. Brown et al. then test whether Form ADV data are associated with prior problems. Because historical Form ADV data are not publicly available, the authors create a measure of operational risk, the ω-score, based on the correlations between contemporaneous Form ADV data and historical hedge fund data. They then test whether the ω-score can predict hedge fund closure, flows, and returns. We also use Form ADV data, but our work differs from Brown et al. in several ways. First, we use historical Form ADV filings to make ex ante predictions of fraud. Second, we focus on fraud rather than their very broad definition of operational risk. (Indeed, of the 126 "problem" hedge fund advisers identified by Brown et al., we find that only six have prior incidents of fraud). Finally, their measure of operational risk includes violations by affiliated firms, such as broker-dealers. These differences are empirically important; we replicate the ω-score of Brown et al. but find it has an insignificant relation with subsequent fraud.

3. Data

3.1. Investment fraud

This study combines two types of data: (1) investment fraud data and (2) disclosures made by investment advisers in their Form ADV filings. To obtain investment fraud data, we search all SEC administrative proceedings and litigation releases[3] that contain the terms "fraud" and "investment adviser" (or "investment advisor") filed from August 2001 through July 2010. From these documents, we identify all cases that involve violations of the anti-fraud provisions in the Investment Advisers Act. Even when another agency initially detects the fraud case, the SEC launches an administrative proceeding, which we observe. The main dependent variable in our paper includes only fraud cases that harm the firm's investment clients. We do not include insider trading, short sale violations, brokerage fraud, or other crimes, unless they cause direct losses to the firm's investment clients.

Many fraud cases span several years and involve multiple legal actions. Fig. 1 shows the timeline of a fraud initiated in September 2002 by K.W. Brown & Co., an investment adviser that traded securities on behalf of clients and for its proprietary account. K.W. Brown & Co. purchased securities but delayed assigning them to specific accounts. Eventually, the firm would

[3]See `www.sec.gov/litigation/admin.shtml` and `www.sec.gov/litigation/litreleases.shtml`.

allocate profitable trades to its proprietary account and unprofitable trades to clients, resulting in losses of over $9 million to the firm's clients. The SEC uncovered problems in March 2003 during a routine examination and notified the firm in June 2003. The fraud continued until March 2004. In April 2005, the firm and its key employees were charged with fraud. The firm and its employees were convicted in December 2007. In January 2008, the SEC filed an administrative proceeding to bar Kevin W. Brown, his wife, and another employee from the securities industry. The firm was deregistered in June 2008.

Because this kind of extended legal scenario is common, and because our goal is to predict fraud rather than detect it, we aggregate all legal actions associated with a single underlying fraud into a single "case" and identify the periods in which fraud occurred. For example, we define the K.W. Brown & Co. fraud case as occurring from September 2002 until March 2004, and use the information from K.W. Brown & Co.'s August 2002 Form ADV filing to predict the initiation of fraud in September 2002. We also use information from K.W. Brown & Co.'s August 2003 Form ADV filing to predict the continuation of fraud into 2004. For the remaining years of the sample, we classify K.W. Brown & Co. as a clean firm. By predicting the occurrence of fraud in 2002-2004, rather than its detection in 2005, we avoid potential biases caused by a correlation between detection and time variation in the predictive variables.

The SEC legal filings include investment fraud cases committed by firms that did not register with the SEC, and thus were not required to file Form ADV (see the next subsection for more detail). To address the economic importance of fraud committed by registered versus non-registered investment advisers, Panel A of Table 1 summarizes fraud cases for both types of firms. Registered firms commit slightly over half of investment fraud cases and are responsible for the overwhelming majority of the dollar losses from fraud. Thus although the scope of our tests is limited to registered investment advisers, these firms are responsible for the most economically meaningful fraud cases.

Panel B summarizes firm-wide fraud, committed with the knowledge of the firm's executive officers, as well as fraud by rogue employees who evade their firms' internal controls. The vast majority of fraud cases are firm-wide. Panel B also summarizes the dollar losses and the duration of the fraud cases. Because fraud often involves the falsification of records, some loss amounts are unavailable, and the available amounts are generally a lower bound, including only the proven losses. Fraud duration is defined as the period extending from the initiation of the fraud until the firm ceases the fraudulent activity. The median fraud case persists for nearly three years. The maximum durations, summarized in the last column of Panel B, reflect the fact that the sample includes cases that were initiated prior to 2001.

To test whether past fraud can predict future fraud, we search SEC administrative proceedings and litigation releases filed from September 1995[4] through July 2001, and create two variables. The first, Past Fraud, is equal to one if a prior administrative proceeding or litigation release shows that the firm has committed fraud. The second, Past Affiliated Fraud, is equal to one if an affiliated firm has committed fraud (affiliation implies the firms are under common control, such as common ownership or executives). Both variables are restricted to include only fraud cases that harmed investment advisory clients. This restriction is consistent with the main dependent variable. Further, Karpoff and Lott (1993), Karpoff et al. (2005), and Murphy et al. (2009) show that firms suffer greater reputational penalties for defrauding counterparties, such as customers, than for defrauding other stakeholders. We match Past Affiliated Fraud to investment advisers using the affiliated firm identifiers from Schedule D of Form ADV. To prevent a look-ahead bias in the predictive regressions, these variables only include fraud cases that had ceased and were publicly revealed before August 1 of that year. For example, in the K.W. Brown & Co. case summarized in Fig. 1, Past Fraud is equal to one only after April 2005 when the fraud had ceased and the first relevant SEC legal filing was publicly accessible.

3.2. Form ADV data

The Investment Advisers Act, which expressly defines and prohibits investment adviser fraud, requires all advisers with more than $25 million in assets under management and with 15 or more U.S. clients to register with the SEC. The Act defines an investment adviser as any entity that receives compensation for managing securities portfolios or providing advice regarding individual securities.[5] Registered investment advisers must file Form ADV to

[4] Online access to administrative proceedings and litigation releases begins in September 1995.

[5] Section 203(b)(3) of the Investment Advisers Act exempts firms with fewer than 15 U.S. clients, that do not advise funds registered under the Investment Company Act, nor "hold themselves out to the public" as investment advisers. Some hedge funds use

disclose past regulatory violations and potential conflicts of interest.

Form ADV contains 12 items and four schedules. Items 1 to 6 contain descriptive information about a firm and its operations. Items 7 and 8 require disclosure of certain conflicts of interest. Item 9 requires disclosure regarding the custody of clients' assets. Item 10 requires disclosure of control persons. Item 11 requires disclosure of past legal and regulatory violations. Item 12 identifies small businesses. Schedules A, B, and C identify the direct and indirect owners of a firm. Schedule D requires disclosure of affiliations with other financial firms.

An SEC website provides a public link to the Investment Adviser Registration Depository, which includes the most recent Form ADV filings from all registered investment advisers.[6] Until recently, investors could access the latest filings only one at a time, and past filings were unavailable. Beginning in January 2010, the SEC began to provide downloadable files of historical Form ADV data.[7] Downloadable files from July 2006 through November 2009 contain summaries of the schedules rather than Form ADV's item data. Downloadable files from December 2009 until the present contain the item data, but not the schedule data.

The SEC provided us with a database of all Form ADV filings from August 2001 through July 2006, including initial filings, amendments, schedules, and the filings of now-defunct firms. These data are not publicly accessible and, to our knowledge, no other researchers have examined them. To create an annual panel for the predictive regressions, we select each firm's most recent filing as of August 1 of each year.[8] This annual panel includes 53,994 firm-year observations representing 13,853 unique investment management firms. We combine the investment fraud documentation and Form ADV data by matching the firms' full legal names.[9]

---

this exemption to avoid registration. A 2004 SEC ruling required hedge fund advisers to register by February 2006, but a U.S. District Court reversed this ruling in June 2006. Despite these exemptions, many hedge fund advisers were registered prior to 2006, either voluntarily or because they also advised other portfolios.

[6]See `www.sec.gov/IARD`.

[7]See `www.sec.gov/foia/docs/invafoia.htm`.

[8]Firms must file Form ADV at least once per year, but often file more frequently; the median firm files 11 times per year. We choose August 1 to maximize the number of annual observations since our set of Form ADV filings ends July 31, 2006.

[9]Of the 251 fraud cases committed by non-registered investment advisers 13 were registered at the time the fraud was initiated, but were deregistered before our sample

3.2.1. Form ADV variables

Table 2 summarizes a cross-section of the investment advisers' characteristics and disclosures, using information from each firm's first Form ADV filing during the sample. Panel A shows that the median firm is wholly employee-owned. Employee Ownership, calculated as in Dimmock et al. (2011), is included because external owners may deter fraud by monitoring employees. The Average Account Size is $55 million, but this variable is highly skewed and the median is only $1.4 million. Percent Client Agents is the percentage of the firm's clients who are agents (e.g., pension fund managers) rather than the direct beneficiaries of the invested funds. On average, 23.2% of a firm's clients are agents. This additional layer of agency is potentially related to fraud because agents have weaker incentives to monitor investment advisers, but may also have greater expertise and financial sophistication. Assets Under Management (AUM) varies greatly across firms. The median AUM is $90 million, but the mean is greater than $2.2 billion.

Panel B of Table 2 tabulates many of the variables disclosed in Form ADV (see the Appendix for detailed definitions). Column one shows summary statistics for the full sample. Column two shows summary statistics for firms in which no fraud is committed from August 1, 2001 through July 2007 (Clean). Column three shows summary statistics for firms in which fraud is committed during the sample period (Fraud). The third column also reports the univariate significance of the difference between clean and fraud firms, using Fisher's exact test.

Item 11 of Form ADV requires each investment adviser to disclose its disciplinary history, as well as that of its (non-clerical) employees, its affiliated firms, and the employees of affiliated firms. The 24 questions in Item 11 are divided into three categories: regulatory, criminal, and civil judicial. From these questions, we create two indicator variables. Past Regulatory equals one if the firm discloses past regulatory violations, indicating sanctions by the SEC, the Commodity Futures Trading Commission, or a self-regulatory organization such as the Financial Industry Regulatory Authority (FINRA). The second variable combines the remaining two categories; Past Civil or Criminal equals one if the firm discloses unfavorable civil judicial decisions related to investment advising, or if the firm discloses criminal convictions. Fraud firms are significantly more likely to report both types of violations.

---

began. We successfully match all fraud cases by currently registered investment advisers.

The disclosure information in Item 11 covers a wide range of regulatory and legal offenses, and the offences are often minor, such as failing to follow protocols for record storage. Minor violations seem to be the norm rather than the exception, and should be interpreted as such: Less than 2.5% of firms that report past violations have a prior instance of fraud. Form ADV does not distinguish whether the investment adviser or its affiliate(s) committed the reported violations, and so there is a strong positive correlation between prior violations and the number of affiliates. To avoid a spurious correlation between the prior violations of affiliated firms and investment adviser fraud, our dependent variables do not include fraud committed by affiliated firms.

Items 7 and 8 of Form ADV require firms to disclose conflicts of interest. From this information we create three variables. Referral Fees equals one if the firm compensates other parties for client referrals. Interest in Transaction equals one if the firm trades directly with its clients or has a direct financial interest in securities recommended to its clients; these practices create potential conflicts and provide a mechanism for fraud. Soft Dollars equals one if the firm directs clients' trades to a brokerage with relatively high commissions and, in return, the broker supplies the adviser with research or other benefits. Since clients pay the costs while the investment adviser realizes the benefits, soft dollars create a potential conflict of interest.

The next four variables are intended to measure monitoring. Broker in Firm equals one if the firm employs registered representatives of a broker-dealer. Trading through an affiliated broker-dealer removes one form of external oversight and provides a mechanism for fraud. Investment Company Act equals one if the firm manages money on behalf of a fund registered under the Investment Company Act, such as a mutual fund. The Act increases regulation and disallows certain conflicts of interest but also indicates the firm's investors are relatively unsophisticated. Custody equals one if the firm has possession, or the authority to obtain possession, of its clients' assets. Custody facilitates fraud by removing external oversight. However, SEC Rule 206(4)-2 requires audits of investment advisers with such custody, including at least one unannounced visit per year, which may reduce the incentive for fraud by increasing the likelihood of detection. Dedicated CCO equals one if the firm's chief compliance officer (CCO) does not have another formal job title. All registered investment firms must designate a CCO who is responsible for ensuring compliance with SEC regulation, but often the CCO has other potentially conflicting roles within the firm.

Hedge Fund Clients equals one if over 75% of the firm's clients are hedge fund clients. We include this variable for two reasons: First, hedge funds are relatively opaque, which could facilitate fraud. Second, prior to 2006 some hedge fund advisers were not required to file Form ADV, which could create a sample selection bias if non-reporting is associated with fraud.

3.3. Fund level data: Returns and fees

To test the relation of fraud risk with performance and fees we require *fund* level data that are not disclosed in Form ADV. We obtain fund level data from the TASS hedge fund, CRSP mutual fund, and PSN Informa databases. We match these databases to the Form ADV sample using firm name, location, and assets under management.[10] For the CRSP mutual fund and PSN Informa databases, we include only equity funds in our sample.

We are able to match 1,511 of the firms in the TASS database (37.2%), which manage 2,848 distinct hedge funds. From TASS we obtain monthly returns, management fees, incentive fees, and other variables. Participation in the TASS database is voluntary and hedge funds are not required to publicly disclose their returns. As a result, the merged hedge funds may not be representative of all hedge funds.

Because all mutual fund advisers must file Form ADV and publicly report their returns, we are able to match all management companies in the CRSP mutual fund database. From this database we obtain monthly returns, expense ratios, and other variables for 2,818 equity funds.

To obtain information on institutional funds we use the PSN Informa database [institutional funds are long-only portfolios managed on behalf of accredited investors, see Busse et al. (2010) for more details]. We are able to match 1,578 of the PSN firms (88.2%), which manage 4,189 distinct portfolios and 89.2% of the aggregate assets under management. Like hedge funds, institutional funds are not required to publicly disclose their returns, and so the funds in our sample may not be representative. In addition to monthly returns, we also obtain information on the posted annual fee charged on a $50 million account (institutional funds can charge clients different fees and the reported fees are only approximate).

In total, 3,123 of the firms in the Form ADV sample match to at least one of the fund return databases, and 314 of the firms match to all three databases. Although the matched firms are only 22.5% of the total Form

[10]For summary statistics and details about matching see the web appendix.

ADV sample, they control the majority of assets under management.

4. Predicting fraud

In this section, we test whether the Form ADV data can predict investment fraud. The purpose of these tests is prediction and, as noted previously, we make no claims regarding causality. Many of the independent variables are endogenous (e.g., a firm's executives may deliberately choose an organizational structure that enables fraud), but because our goal is prediction rather than establishing causality, the potential endogeneity of the independent variables does not change our interpretation.

A major caveat in interpreting our findings is that we observe only detected fraud. Three factors affect observed fraud: the unobservable true rate of fraud, the probability of detection given a fixed level of monitoring, and the allocation of monitoring resources. Ideally, the regressions will predict the true rate of fraud. However, if certain predictive variables are correlated with either monitoring or detection, this relation could affect the interpretation of the results. Further, the predictive variables could be correlated with monitoring and detection for two reasons. First, any predictive variable that decreases the probability of detection will increase the incentive to commit fraud. In general, this problem biases against significant results because predictive variables that are associated with a higher rate of fraud will also be associated with a lower detection rate. Second, if the difficulty of detecting fraud affects the allocation of monitoring resources, this may, or may not, outweigh the added difficulty of detecting fraud. These difficulties could cause the empirically observed relations to differ from the actual relation between firms' disclosures and the unobservable true rate of fraud.

We address the issue of undetected fraud in two ways. First, we conduct extensive out-of-sample tests to ensure the predictions are robust. Second, although the panel of independent variables ends in 2006, we search for detected fraud cases through July 2010. For each case, we identify when the fraud occurred. In the predictive regressions, the dependent variable is the occurrence of fraud in a given year, even if the fraud remains undetected for years. Unfortunately, a direct test of the relation between these variables and fraud detection is not possible. Certain types of fraud may go undetected, a possibility that could bias the results.

4.1. Prediction models

Panel A of Table 3 shows the results of probit regressions that predict investment fraud using Form ADV disclosures. In column one, the sample is a cross-section of firms. The independent variables are taken from each firm's first Form ADV filing during the sample period; the dependent variable equals one if the firm commits fraud *at any time* between its first filing during the sample period and July 2007. This specification includes indicator variables for the year in which the firm first filed Form ADV. In this cross-sectional specification, the z-scores are based on robust standard errors.

In the remaining columns, the sample is an unbalanced panel of firm-year observations. The dependent variable equals one if a fraud occurs during the subsequent 12 months. In columns two and three, the sample includes all firm-year observations. In column four, the sample excludes firms with a history of fraud identified in prior SEC administrative proceeding or litigation releases. In the last column, the sample also excludes firms that disclose the relatively minor legal or regulatory violations in Item 11, committed by either the firm itself or an affiliated firm. For these panel specifications in columns two through five, the z-scores are based on standard errors clustered by firm and year. The chi-square tests at the bottom of each column show the significance of the overall model.

Past Fraud is insignificant in both the cross-sectional and panel regression. There are few firm-year observations with prior fraud because many firms that commit fraud subsequently cease operations, and so the regressions have low power with respect to this variable.[11] Past Affiliated Fraud does not predict fraud in the cross-sectional regression, but in one of the panel regressions, the coefficient is marginally significant and negative. Unlike the other predictive variables, Past Fraud and Past Affiliated Fraud are not disclosed in Form ADV.

Past Regulatory and Past Civil or Criminal are both significant positive predictors of subsequent fraud, even in the sample that excludes firms with prior fraud. The simplest explanation is that past problems, although frequently minor, indicate poor internal controls or unethical management. But two additional explanations exist: Past violations could increase the rate of detected fraud due to the increased probability of an SEC examination. Also, because each firm must disclose both its own prior violations and those of

[11]See web appendix Table E for information on firm survival following fraud.

its affiliated firms, prior violations are strongly correlated with the size and scope of an investment firm's affiliated businesses (i.e., financial conglomerates are more likely to report prior violations). These affiliations could increase conflicts of interest and provide the means to commit fraud.

The next three variables measure several potential conflicts of interest between investment advisers and their clients. Referral Fees has a significant positive relation with subsequent fraud. Fraud firms could be relatively willing to pay referral fees because fraud increases the marginal profit per dollar managed. Interest in Transaction also has a significant positive relation with subsequent fraud. When investment managers take the opposite side of a transaction from their clients, this arrangement creates an obvious conflict of interest and also provides a mechanism for fraud. Soft Dollars does not significantly predict fraud.

We include several variables to measure the monitoring of investment advisers. Broker in Firm has a significant positive relation with subsequent fraud. Trading through an in-house brokerage removes external oversight and creates a mechanism for committing certain types of fraud. Investment Company Act has a significant positive relation with subsequent fraud. The Act increases regulatory oversight of these firms, which could increase the probability that fraud is detected. Alternatively, the true rate of fraud could be higher because these firms exploit their clients' lack of financial sophistication. The next three variables, Custody, Dedicated CCO, and Majority Employee Owned, are not significant in the panel regressions, although Custody and Dedicated CCO are significant in the cross-sectional regression. Note that even if a variable is insignificant in these regressions, its disclosure may still have a beneficial deterrence effect.

The next three variables (Logarithm of Average Account Size, Percent Client Agents, and Hedge Fund Clients) also measure monitoring but are based on client characteristics. Although all clients have an incentive to monitor, large investors have a stronger incentive and possibly a greater ability to do so. The results for the Logarithm of Average Account Size show that larger investors are associated with fewer subsequent fraud cases. This result could be a selection effect, meaning that large investors select honest managers. Alternatively, because of financial sophistication or economies of scale in monitoring, large investors may deter fraud because of a higher probability of detection. Both arguments suggest that large investors are associated with a lower rate of fraud rather than a lower detection rate.

Percent Client Agents, the second variable measuring client characteristics, has a significant positive relation with subsequent fraud. After conditioning on average account size, firms whose clients include a high proportion of agents are more likely to commit fraud. Although agents may have reputational concerns and greater financial sophistication, they do not bear the full cost of fraud, which reduces their incentive to monitor and suggests that they can be swayed through gifts or kickbacks. The reduced incentives of agents appear to outweigh their potentially higher sophistication.

Hedge Fund Clients is an indicator for firms that primarily manage hedge funds. The results do not provide evidence of a relation between hedge fund management and fraud. Hedge funds are relatively non-transparent, however, and so the detected fraud cases may understate the true frequency of fraud that occurs within hedge funds. Moreover, not all hedge funds were required to file Form ADV during the early part of the sample, which could create a sample selection bias. Nonetheless, in annual cross-sectional regressions (Table 4) we find that the coefficient on hedge fund management is not significantly different in the later years of the sample, which suggests that sample selection is not a problem.

4.2. The economic interpretation of the prediction models

The probit regressions in Panel A of Table 3 show that the Form ADV variables have a statistically significant relation with subsequent fraud. This finding is important, but the key question of interest is whether the overall model would enable an investor to avoid fraud. To address this question, we take the predicted values from the regressions and examine the tradeoff between correctly predicted fraud cases and the false positive rate. False positives, which occur when the model incorrectly predicts that a clean firm will commit fraud in the subsequent year, can be interpreted as the opportunity cost to investors of erroneously limiting their investment opportunity set. Although failing to predict fraud is likely more costly than mistakenly avoiding an honest investment adviser, an investor would need to avoid multiple honest advisers for every fraud avoided. We address the costs of false positives in Section 5.

To illustrate the possible tradeoffs between false positives and predicted fraud, Fig. 2 shows a receiver operating characteristic (ROC) curve for the prediction model in the second column of Panel A of Table 3. The points on the ROC curve are generated non-parametrically by taking each observation's predicted value from the probit model as a cut-point, and then computing both the proportion of fraud firm-years correctly predicted and the false positives. Random prediction of fraud would result in a straight 45-degree line. Initially the curve rises steeply, showing a considerable number of fraud firm-years could be avoided at a low false positive rate.

The ROC curve in Fig. 2 shows the full range of all possible tradeoffs between the prediction of fraud and false positives. Following a similar format as Dechow et al. (2011), in Panel B of Table 3 we provide greater detail for one possible tradeoff, the proportion of fraud firm-years that could be predicted within sample at a false positive rate of 5%. The columns in Panel B correspond to the columns in Panel A. For example, the model in the second column correctly predicts 150 of 517 fraud firm-years (29.0%) at a false positive rate of 5% (we incorrectly predict fraud in 2,673 clean firm-years that are associated with 885 distinct firms). The last row of Panel B shows the percentage of total dollar losses that could have been avoided at a false positive rate of 5%. The dollar losses from fraud are winsorized at the 99^{th} percentile because of several extreme outliers (e.g., Madoff's $18 billion Ponzi scheme). For multiyear fraud cases, we distribute the losses evenly across years. The model in the second column correctly predicts 41.3% of the total dollar losses from fraud at a false positive rate of 5%, which indicates that the model predicts economically meaningful fraud cases and not merely small cases.

The results in Panel B are similar for all models, except for the specification reported in the last column, in which the subsample does not include firms that report prior legal or regulatory violations, either by the firm or its affiliates. For this subsample, both the percentage of fraud firm-years predicted and the percentage of the total dollar losses to fraud avoided are substantially lower. By comparison, the results for the subsample that excludes firms with prior publicly revealed fraud, shown in the fourth column, are very similar to the full sample. Thus, the difference in the last column is not due to some firms committing fraud numerous times. Rather, fraud is relatively easy to predict among firms with past regulatory and legal violations.

4.3. Out-of-sample prediction of fraud

A key concern for any prediction model is out-of-sample validity. In this subsection, we test whether the within sample predictions, reported in Panel B of Table 3, are robust out-of-sample. We do this in two ways: Panel C summarizes the out-of-sample predictive performance of each model in the post-2007 period. Panel D shows the results from K-fold cross-validation tests, which are explained in the next subsection.

The observations used in the prediction models reported in Panel A of Table 3 include only firm-years prior to August 1, 2007. To conduct an out-of-sample test, we search the SEC administrative proceedings and litigation releases and identify fraud cases that occurred between August 1, 2007 and July 31, 2010. Using Form ADV filings as of August 1, 2006 we assign each firm a predicted value based on the coefficients estimated within sample. We then test whether these predicted values can accurately classify the out-of-sample fraud risk of the firms.

Panel C of Table 3 shows the proportion of fraud cases correctly predicted at a false positive rate of 5%. The proportion of fraud cases predicted out-of-sample is usually higher than within sample, although given the small number of observations this difference is not statistically significant. Also, although we use the within-sample cutoff values to classify the firms out-of-sample, the false positive rate does not increase.

4.3.1. K-fold cross-validation tests

As a further robustness test of the predictive models in Panel A of Table 3, we perform K-fold cross-validation tests over the period August 2001 through August 2007. The idea behind these tests is simple. Each model is estimated on a randomly selected subsample of firms, and the coefficient estimates from this subsample are used to classify the firms in the hold-out sample. Specifically, each firm in the sample is randomly assigned to one of 10 groups (note that we randomly assign *firms*, and not *firm-years*, to avoid overstating the results due to non-independence). 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 hold-out sample. We repeat this process 20 times, for a total of 200 hold-out samples.

The results, shown in Panel D of Table 3, indicate that the predictive power of the models is only slightly lower in the hold-out samples. For example, the specification in the second column correctly predicts 150 fraud firm-years within sample, compared to an average of 143.3 fraud firm-years in the hold-out samples. The K-fold test predicts a minimum of 135 and a maximum of 149 fraud firm-years across 20 repetitions, which suggests the model is quite stable.

The results of the out-of-sample and K-fold cross-validation tests support the robustness of the fraud predictions in Panel B. Note that these are robustness tests of the models' overall predictions and do not provide evidence as to the robustness of the individual coefficients. Overall, the results from the four panels of Table 3 show that the information investment advisers are required to disclose is relevant and useful for predicting fraud.

4.4. Annual cross-sectional regressions

Although the models presented in Table 3 use observations from the entire sample period, which allows for relatively powerful tests, this may obscure time effects, which could arise in several ways. First, the actual rate of fraud could change over time due to changes in the legal or operating environment (e.g., poor performance could decrease the benefits of a reputation for honesty, thus increasing the incentive to commit fraud). Second, the detection rate could change over time. The median fraud persists for five and a half years until detection. This suggests that the dependent variable for the 2006 cross-section likely includes fewer than half of the fraud cases that actually occurred in that year. By contrast, the 2001 cross-section likely includes a much higher proportion of the fraud cases that actually occurred in that year.

To examine whether there are time effects in the prediction of fraud, Panel A of Table 4 shows annual, cross-sectional probit regressions that predict investment fraud that occurs during the subsequent 12 months. For example, the model in column one uses Form ADV data available on August 1, 2001 to predict fraud that occurs from August 2001 through July 2002. Because fraud cases can persist for multiple years, these annual regressions are not independent, and aggregating coefficients across years could lead to faulty conclusions.

We test whether the coefficient estimates are significantly different across years with Wald tests. Because the same firm can appear in multiple years, we adjust the Wald tests for non-independence. The coefficients for Dedicated CCO and for Investment Company Act are significantly different across years at the 1% and 10% levels, respectively. Both variables are significant only in the early years of the sample. This change is partially due to the mutual fund late trading scandal that occurred in the early years of the sample. The firms involved in these cases managed funds that were registered under the Investment Company Act, and were mostly large financial conglomerates, which are more likely to have a dedicated CCO. Custody and Majority Employee Owned are also significantly different across years.

Note that because there are fewer observations in these annual cross-sectional regressions, the Wald tests have low power to reject the hypothesis that the coefficients are equal across years. For example, Referral Fees is significant in Table 3, but in Table 4 Referral Fees is significant in only two years. We cannot reject that the coefficients are jointly equal to zero, nor that they are jointly equal to the full sample coefficient.

Panel B of Table 4 shows the ability of the cross-sectional regressions to predict fraud within sample at a 5% false positive rate for each year. Chi-square tests show that the prediction rate of the annual cross-sectional regressions is significant in each year, which suggests the results in Table 3 are not driven by a single period. The proportion of fraud cases predicted, however, is lower in the last three years of the sample. This is partially due to the mutual fund late trading cases that occurred in the earlier years of the sample. Even after removing these cases, however, predictive accuracy appears to decline. This decline could indicate that the actual predictability of fraud has declined over time; although given the relatively low number of fraud cases in some years and the power of the annual tests, we cannot draw strong conclusions.

4.5. Initiation versus continuance of fraud

In the previous tests, the dependent variable does not distinguish between the initiation of a new fraud and the continuance of a preexistent fraud. In this section, by contrast, we test whether Form ADV data can be used to predict the initiation of a new fraud. This is important for three reasons. First, predicting fraud prior to initiation likely minimizes the harm. Second, ongoing fraud could affect the predictive variables, and thus the previous tests blur the distinction between predicting future acts of fraud versus detecting ongoing fraud. Third, initiating a new fraud and continuing a preexistent fraud are economically different decisions. Lott (1996) shows that firms may initiate fraud in response to changes in their cost structure. Dechow et al. (1996) and Dechow et al. (2011) find evidence that companies initiate accounting fraud in response to corporate performance. Thus, certain predictive variables might measure a time-varying factor that triggers the initiation of fraud, whereas other variables might measure a time-invariant propensity toward fraud.

To address these issues, Panel A of Table 5 shows the results of a sequential logit regression with standard errors clustered by firm. The equation displayed in column one predicts the initiation of fraud in the subsequent year. The results in this column are qualitatively similar to the results in Table 3, and the chi-square test shows that the overall equation is statistically significant. The equation displayed in the second column predicts whether firms that have previously initiated a fraud will continue the fraud in subsequent year. The insignificant chi-square test shows that the Form ADV variables have no incremental ability to distinguish which initiated fraud cases are continued into subsequent years. The insignificance of the second equation likely reflects that the Form ADV variables are quite stable over time and suggests that these variables primarily measure a time-invariant component of the propensity for fraud.

Panel B of Table 5 shows the proportion of initiated fraud cases that could be predicted within sample.[12] At a false positive rate of 5%, the model predicts 37.9% of initiated fraud cases, which suggests that the predictive ability of the model is not entirely due to continued fraud cases.

4.6. Firm-wide fraud versus fraud by rogue employees

In some cases, the executives of the firm commit or are aware of the fraud. In other cases, rogue employees evade their firms' internal control systems. A potential concern with the prior tests is that rogue employee fraud is likely more frequent at firms with many employees. If the predictive variables are correlated with the number of employees, this could lead to spurious correlations with rogue employee fraud. To ensure that this problem does not drive the results, we compare the predictability of firm-wide and rogue employee fraud.

Panel A of Table 6 shows the results of a multinomial probit regression. In the first column, the dependent variable equals one if the firm commits a firm-wide fraud in the subsequent year. In the second column, the dependent variable equals one if a rogue employee commits fraud in the subsequent year. The results for firm-wide fraud are very similar to the full sample results, reported in column 3 of Table 3. At a false positive rate of 5% the model can predict 24.1% of all firm-wide fraud cases, compared to 29.7% of all fraud cases. Panel B shows that, at a 5% false positive rate, the model can predict 74.6% of rogue employee fraud cases. Thus, although firm-wide fraud is more difficult to predict, the results in Table 3 are not entirely due to rogue employee fraud.

[12] We do not show prediction results for the proportion of preexistent fraud cases that are continued into the subsequent year. These predictions could be made only if preexistent fraud cases were readily observable, which is not the case.

5. Are investors compensated for fraud risk?

Given that fraud is predictable, why would investors allocate money to firms with high fraud risk? One possibility is that the characteristics associated with high fraud risk may provide offsetting benefits that improve investment performance. Some investors could voluntarily accept high fraud risk in return for higher expected returns. Another possibility, which follows from the theoretical model of Lott (1996), is that investors pay a premium for advisers with low fraud risk, and pay lower fees to advisers with high fraud risk. In this section, we test whether investors are compensated for bearing fraud risk through higher returns or lower fees.

Until this point, the unit of observations has been firm-years. To test whether investors are compensated for fraud risk, however, requires fund data, which are only available for a subset of investment advisers. Hedge fund and institutional fund advisers are not required to report return data, which could create a selection bias if fraudulent advisers choose not to report returns. Thus, the results in this section are less general and may not apply to the full sample. Data are available for all mutual funds and mutual fund advisers, however, and so there is no selection bias for this category of funds.

We measure each firm-year observation's fraud risk as the predicted value from the probit regression in the third column of Table 3. In August of each year, we assign each fund its adviser's predicted fraud risk. We classify funds as high fraud risk if they are advised by a firm whose predicted fraud risk is greater than the 95^{th} percentile of clean firms. This classification corresponds to the prediction cutoff used in Panel B of Table 3.

For each of the three return database samples, we form two equally weighted portfolios. In August of each year, we assign all funds classified as high fraud risk to one portfolio. The remaining funds are assigned to the low fraud risk portfolio. We then find the equally weighted returns of the portfolios for the subsequent 12 months. See the web appendix for additional details and for results using value weighted portfolios.

In Panel A of Table 7 we estimate alphas for the TASS hedge fund sample using the Fung and Hsieh (2001) eight factor model.[13] For the CRSP mutual funds and PSN institutional funds we estimate alphas using the Carhart (1997) model. For all three of the return database samples, we estimate

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

alphas over the 72 month period from August 2001 to July 2007. Within all three return database samples, we do not find evidence that fraud risk is associated with higher risk adjusted returns.

The regressions in Panel B of Table 7 test whether investors receive compensation for predicted fraud risk through lower fees. The control variables follow Cassar and Gerakos (2010) for hedge funds and Khorana et al. (2009) for mutual funds and institutional funds. All specifications control for fund style,[14] fund size, fund management company size, and year fixed effects. Within each of the three return database samples, we estimate pooled regressions with standard errors clustered by firm. We do not find evidence that high fraud risk is associated with lower fees in any of the samples.

The results in Panels A and B of Table 7 do not provide evidence that investors receive compensation for fraud risk. This finding has important implications for interpreting the prediction results. In Section 4.2, we noted that using a fraud prediction model requires an investor to avoid a considerable number of honest advisers (false positives). The results in this section suggest that there is little cost to avoiding these false positives.

The lack of compensation for fraud risk does not necessarily imply that investors are irrational. First, investors could be compensated in some other, unmeasured fashion. Second, our measure of predicted fraud risk uses Form ADV disclosures. Investors may predict fraud risk based on reputations, personal contacts, or other information; investors could be compensated for some alternative measure of fraud risk. Finally, because of barriers to accessing and using the Form ADV disclosures, investors might perceive the costs of estimating fraud risk to outweigh the benefits.

6. Data access and implementation

In this section, for each year in the sample we test how well investors could have implemented predictive models using only Form ADV data that had previously been publicly accessible. These tests differ from the predictive regressions in Table 3, which use information from the full sample period and thus do not directly address how well an investor could have predicted fraud

[14] We measure fund style using the Primary Category variable for the TASS hedge funds, Lipper categories for the CRSP mutual funds, and reported market capitalization and value/growth categories for the PSN funds.

during the sample (e.g., the fraud predictions in 2003 are based on coefficients estimated using data from 2001 through 2006).

During the sample period, the SEC did not provide public access to historical Form ADV filings; investors could access only a contemporaneous cross-section. For this reason, we compare two types of predictive models. In the first, we estimate predictive models that use only the contemporaneous cross-section of Form ADV filings. These tests mimic the predictions an investor could have made during the sample period, given the actual data access policies in place. In the second, we estimate predictive models that use data from an annual panel of historical Form ADV filings. These tests mimic the predictions an investor could have made if historical filings had been publicly accessible.

Table 8 shows the results of fraud prediction models that use only the contemporaneously accessible cross-section of Form ADV filings as of August 1 of each year. To illustrate, in the column labeled Aug '05, the independent variables are taken from each firm's most recent Form ADV filing as of August 1, 2005. The dependent variable equals one for all firms with an observable prior fraud case (i.e., a fraud case that occurred between September 28, 1995 and July 31, 2005, and which was identified in an SEC administrative proceeding or litigation release filed before July 31, 2005). We use the coefficient estimates from this regression to make out-of-sample predictions of the fraud cases that occur between August 1, 2005 and July 31, 2006.

For comparison purposes, Table 9 shows the results of fraud prediction models that use an annual panel of historical Form ADV filings. Like Table 8, these regressions use only information that existed at the time of the prediction. But unlike Table 8, they use information that was not contemporaneously accessible by the public. To illustrate, in the column labeled Aug '05, the independent variables are taken from each firm's Form ADV filings as of August 1 in 2001, 2002, 2003, and 2004. For each August 1 firm-year observation, the dependent variable equals one if the firm commits fraud during the subsequent 12 months, and the fraud is publicly revealed before August 1, 2005. We combine the coefficient estimates from this model with each firm's Form ADV data as of August 1, 2005 to make out-of-sample predictions of the fraud cases that occur between August 1, 2005 and August 1, 2006.

The models presented in Tables 8 and 9 differ in several ways. Most obviously, the panel models in Table 9 use more data to estimate fraud risk than do the cross-sectional models in Table 8. More important, the models in Table 8 are backward looking: they show the relation between contemporaneous variables and *past* fraud. If the contemporaneous Form ADV filings are all that is publicly accessible, then investors can only estimate fraud risk from backward-looking regressions [e.g., Brown et al. (2008)]. If historical filings are accessible, investors can estimate forward-looking prediction models and then estimate fraud risk by combining the estimated coefficients with the contemporaneous disclosures of the firms. This is a conceptually important distinction. The backward-looking regressions only include the subsample of firms that survived the legal consequences of committing fraud, which could bias the coefficients.[15] The forward-looking models in Table 9 have one disadvantage, however: These models require at least two years of data to estimate, and so it is not possible to estimate this model for the year beginning August 1, 2001.

For the regressions reported in Tables 8 and 9 our main interest is not the coefficient estimates, but rather the models' ability to predict fraud. Specifically we are interested in two issues. First, could the prediction results in Table 3 have been achieved during the sample. Second, we are interested in comparing the predictive ability of the models in Tables 8 and 9. Between August 1, 2002 and August 1, 2007 there were a total of 413 fraud firm-years. At a false positive rate of 5%, the cross-sectional regressions, shown in Table 8, predict 24.7% of the fraud cases (the fraud firm-years predicted between August 2002 and August 2007 sum to 102). The panel regressions using all prior years, shown in Table 9, predict 31.2% of the fraud firm-years (a sum of 129 fraud firm-years). A chi-square test of classification accuracy shows that the panel regressions in Table 9 predict a significantly larger number of fraud cases (p-value<0.01). Although the absolute difference in predictive accuracy is only 6.5 percentage points (an improvement of 26.3% relative to the model in Table 8), these tests provide evidence that public access to historical Form ADV filings could benefit investors. Moreover, the marginal cost to the SEC of allowing public access would be quite low.

To implement a fraud prediction model, such as those tested in this paper, an investor would have had to collect manually a large number of Form ADV filings, convert the filings into a database and estimate a prediction model.

[15]Nearly a third of the firms at which fraud is detected cease operations within one year. Duration models, presented in the web appendix, show a strong relation between the detection of fraud and firm closure.

For most investors, the cost of individually downloading thousands of Form ADV filings may well have exceeded the perceived benefits. This problem is exacerbated by the fact that investors are atomistic: Even if the aggregate benefit of processing the disclosed information outweighs the cost to a single investor, the benefit to any single investor may be insufficient. As shown by Becker (1968), the socially optimal level of a crime occurs when the marginal benefit from a further reduction in the crime is equal to the marginal cost of enforcement. Allowing public access to historical Form ADV filings would reduce the marginal cost of increased enforcement by facilitating investors' use of these data. This, in turn, should reduce the marginal benefit to an investment adviser of committing fraud due to an increase in the probability of detection. Thus, improved public access to these disclosure data could reduce the occurrence of fraud.

7. Conclusion

This paper finds that required disclosures related to past regulatory and legal violations, conflicts of interest, and monitoring are significant predictors of investment fraud. We stress, however, that prediction does not imply a causal relation between the disclosed information and fraud. Although fraud is predictable, we do not find evidence that investors are compensated for fraud risk. To explain this puzzling fact, we examine the barriers to implementing fraud prediction models.

If, during the period August 2001 through August 2007 investors had avoided the 5% of firms with the highest ex ante predicted fraud risk, they could have avoided more than $4 billion in losses from fraud. 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 period the direct costs of disclosure were at most $500 million. Thus, even ignoring the deterrent effect disclosure, this simple, back-of-the-envelope calculation suggests that the benefits of Form ADV substantially outweigh the costs. During the sample period, the investing public's ability to develop and use predictive models based on Form ADV data was potentially limited because the SEC did not provide access to historical data. As a result, the realized benefits of disclosure during the period may have been lower. The results suggest that improving public access to comprehensive historical disclosures could increase the benefits these disclosures were meant to provide.

Table 1
Summary of investment fraud

This table summarizes cases of investment fraud committed by investment advisers between August 2001 and July 2010 as reported in SEC administrative proceedings and litigation releases. Registered denotes firms that file a Form ADV with the SEC. Firm-wide fraud is committed by high level executives, or at the very least, with the firms' implicit acceptance. Rogue employee fraud is committed by individuals who evade their firms' internal control systems and the firms do not knowingly benefit.

Panel A: Registered vs. non-registered advisers

	Total	Firm-wide	Rogue employee	Investor losses ($ billion)
Non-registered	251	244	7	4.5
Registered	258	217	41	32.4
Total	509	461	48	36.9

Panel B: Fraud characteristics

		Investor losses ($ million)				Duration (years)		
	Obs.	Mean	Median	Max	Missing obs.	Mean	Median	Max
Firm-wide	217	196.3	6.0	18,000.0	56	4.0	3.1	20.8
Rogue employee	41	25.4	3.0	300.0	8	3.2	2.5	11.1
Total	258	167.2	5.1	18,000.0	64	3.9	2.9	20.8

Table 2
Summary of investment advisory firms

This table summarizes information from each firm's first Form ADV filing during the period August 2001 through July 2006. There are 13,853 unique firms in the sample. 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. Past Fraud equals one if the firm is identified as committing fraud in a previous SEC filing. Past Affiliated Fraud equals one if the firm's affiliates have been identified as committing fraud in a previous SEC filing. Past Regulatory equals one if the firm reports past regulatory violations. Past Civil or Criminal equals one if the firm reports past civil or criminal violations. Referral Fees equals one if the firm compensates any party for client referrals. Interest in Transaction 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. Investment Company Act equals one if the firm is registered under the Investment Company Act of 1940. Custody equals one if the firm has custody of clients' cash or securities. Dedicated CCO equals one if the chief compliance officer has no other job title. Hedge Fund Clients equals one if more than 75% of the firm's clients are hedge funds. The column Clean (Fraud) summarizes firms in which a fraud is not committed from first filing through July 2007 (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	25^{th}	50^{th}	75^{th}
Employee Ownership	68.2%	44.2	0.0	100.0	100.0
Avg. Acct. Size ($ thousand)	55,361	328,522	339	1,442	21,667
Percent Client Agents	23.2%	32.6	0.0	8.3	30.0
Assets Under Mgmt. ($ million)	2,213	16,433	37	90	400
Firm Age (years)	5.1	7.7	0.4	1.1	8.1

Panel B: Firm disclosures			
	All	Clean	Fraud
Past Fraud	0.2%	0.2	1.6***
Past Affiliated Fraud	1.6%	1.6	2.6
Past Regulatory	12.1%	11.9	32.6***
Past Civil or Criminal	3.3%	3.1	12.5***
Referral Fees	40.0%	39.7	59.8***
Interest in Transaction	30.4%	30.1	52.2***
Soft Dollars	55.7%	55.6	63.0**
Broker in Firm	40.8%	40.4	66.3***
Investment Company Act	9.8%	9.6	29.0***
Custody	23.9%	23.7	33.7***
Dedicated CCO	10.7%	10.7	12.4
Hedge Fund Clients	13.4%	13.5	6.2***

Table 3
Predicting fraud

The full sample consists of 53,994 firm-year observations. In the first column, the sample includes only each firm's first Form ADV filed during the sample period. In the remaining columns the independent variables are taken from each firm's Form ADV filing as of August 1 each year from 2001 through 2006. In the second and third columns, the full sample is included. In the fourth column, the sample excludes firms with a previously disclosed fraud. In the fifth column, the sample excludes all firms that disclose in Item 11 of Form ADV any type of prior legal or regulatory violation, either by the firm itself or an affiliated firm. Column one of Panel A shows the results of a cross-sectional probit regression predicting fraud. The dependent variable equals one if the firm commits fraud in any subsequent year through July 2007. Standard errors are robust. Columns two through five show the results of pooled probit regressions predicting fraud. The dependent variable equals one if the firm commits fraud in the subsequent year. Standard errors are clustered by firm and year. In the interest of brevity the constants are not reported. 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 fraud that could be predicted within sample. Panel C shows the out-of-sample performance of each model, using Form ADV filings in August 2006 to predict fraud cases that occur between August 2007 through July 2010. Panel D shows the results from K-fold cross-validation tests.

Panel A: Predictors of fraud					
	Cross section	Full sample	Full sample	No prior fraud	No prior violations
Past Fraud	0.329		0.272		
	[0.98]		[1.46]		
Past Affiliated Fraud	-0.224		-0.184	-0.196*	
	[1.14]		[1.54]	[1.68]	
Past Regulatory	0.175**	0.284***	0.282***	0.285***	
	[2.25]	[4.20]	[4.16]	[4.15]	
Past Civil or Criminal	0.223*	0.191**	0.200**	0.209**	
	[1.93]	[2.13]	[2.29]	[2.32]	
Referral Fees	0.135**	0.100*	0.099*	0.099*	0.139**
	[2.09]	[1.79]	[1.79]	[1.78]	[2.40]
Interest in Transaction	0.138*	0.197***	0.198***	0.196***	0.184**
	[1.93]	[2.89]	[2.91]	[2.86]	[2.24]
Soft Dollars	-0.029	-0.051	-0.046	-0.040	-0.073
	[0.46]	[0.89]	[0.81]	[0.71]	[1.10]
Broker in Firm	0.237***	0.118**	0.120**	0.120**	0.096
	[3.89]	[2.01]	[2.05]	[2.02]	[1.55]
Investment Co. Act	0.103	0.263***	0.269***	0.278***	0.273***
	[1.39]	[3.29]	[3.36]	[3.58]	[2.83]
Custody	0.309***	0.094	0.097	0.088	0.028
	[3.79]	[1.43]	[1.50]	[1.36]	[0.33]
Dedicated CCO	0.288***	-0.088	-0.085	-0.085	-0.056
	[2.67]	[0.86]	[0.82]	[0.82]	[0.53]
Majority Emp. Owned	0.045	0.009	0.001	0.008	0.033
	[0.65]	[0.11]	[0.02]	[0.10]	[0.37]
Log (Avg. Acct. Size)	-0.043***	-0.072***	-0.070***	-0.065***	-0.028
	[3.45]	[4.25]	[4.12]	[3.68]	[1.12]
Percent Client Agents	0.001	0.003***	0.003***	0.003***	0.003***
	[1.40]	[3.91]	[3.88]	[3.77]	[2.91]
Hedge Fund Clients	-0.035	0.031	0.031	0.020	0.030
	[0.27]	[0.27]	[0.27]	[0.18]	[0.22]
Log (AUM)	0.036***	0.060***	0.059***	0.054***	0.020
	[3.76]	[4.10]	[3.98]	[3.57]	[0.93]
Log (Firm Age)	0.014	0.002	0.002	0.002	0.008
	[1.18]	[0.20]	[0.19]	[0.20]	[0.66]
Model chi-square	175.2***	181.5***	198.9***	176.9***	63.2***
Observations	13,853	53,994	53,994	53,750	45,920

Panel B: Within sample predictions					
# Fraud	193	517	517	501	310
Fraud predicted	59	150	152	140	44
	30.6%	29.0	29.4	27.9	14.2
# Clean firms	13,660	53,477	53,477	53,249	45,610
Clean firm false positives	683	2,673	2,673	2,662	2,280
	5.0%	5.0	5.0	5.0	5.0
Prop. of $ losses avoided	37.4%	41.3	43.0	40.5	7.9

Panel C: Out-of-sample predictions (August 2007 - July 2010)					
	Cross section	Full sample	Full sample	No prior fraud	No prior violations
# Fraud	27	27	27	25	18
Fraud predicted	9	10	9	7	1
	33.3%	37.0	33.3	28.0	5.6
# Clean firms	10,356	10,356	10,356	10,293	8,912
Clean firm false positives	517	517	517	514	445
	5.0%	5.0	5.0	5.0	5.0
Panel D: K-fold cross-validation hold out sample predictions (August 2001 - July 2007)					
Avg # fraud predicted	53.6	143.3	142.4	129.7	35.0
Avg % fraud predicted	27.8%	27.7	27.5	25.9	11.3
Stdev fraud predicted (#)	1.39	3.64	3.75	4.32	2.66
Min # fraud predicted	51	135	133	120	32
Max # fraud predicted	56	149	148	137	42
Avg # false positives	678.4	2,669.2	2,669.2	2,658.2	2,275.8
Avg % false positives	5.0%	5.0	5.0	5.0	5.0
Stdev false positives	0.68	0.95	0.95	0.99	0.91
Min # false positives	677	2,668	2,668	2,656	2,274
Max # false positives	679	2,671	2,671	2,660	2,277

Table 4

Annual cross-sectional regressions

The sample consists of 53,994 firm-year observations. Each column contains an annual cross-sectional regression, in which the independent variables are taken from each firm's Form ADV filing as of August 1 each year from 2001 through 2006. Panel A shows the results of annual cross-sectional probit regressions predicting fraud. The dependent variable equals one if the firm commits fraud in the subsequent 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. Panel B shows the proportion of fraud that could be predicted within sample.

Panel A: Predictors of fraud						
	Aug '01	Aug '02	Aug '03	Aug '04	Aug '05	Aug '06
Past Fraud	0.280	0.172	-0.337	0.451	0.625**	0.553*
	[0.82]	[0.52]	[0.78]	[1.49]	[2.25]	[1.81]
Past Affiliated Fraud	-0.321	-0.241	-0.194	-0.010	0.201	-0.051
	[1.23]	[1.15]	[1.11]	[0.05]	[0.94]	[0.19]
Past Regulatory	0.187*	0.235**	0.350***	0.371***	0.248**	0.050
	[1.89]	[2.42]	[3.79]	[3.44]	[2.01]	[0.37]
Past Civil or Criminal	0.239*	0.212*	0.092	0.328**	-0.095	0.342*
	[1.73]	[1.66]	[0.71]	[2.41]	[0.47]	[1.76]
Referral Fees	0.041	0.021	0.071	0.130	0.187*	0.255**
	[0.45]	[0.26]	[0.85]	[1.51]	[1.90]	[2.25]
Interest in Transaction	0.265***	0.252***	0.235**	0.138	0.124	-0.029
	[2.78]	[2.63]	[2.48]	[1.23]	[1.03]	[0.23]
Soft Dollars	-0.037	-0.067	-0.015	-0.020	-0.075	-0.190
	[0.40]	[0.78]	[0.17]	[0.21]	[0.74]	[1.53]
Broker in Firm	0.202**	0.127	0.071	0.043	0.006	0.147
	[2.33]	[1.51]	[0.84]	[0.45]	[0.05]	[1.23]
Investment Co. Act	0.245**	0.325***	0.306***	0.264**	-0.120	-0.081
	[2.43]	[3.25]	[2.89]	[2.26]	[0.73]	[0.42]
Custody	0.006	0.084	0.166*	0.061	0.246**	0.339***
	[0.06]	[0.89]	[1.76]	[0.58]	[2.37]	[2.65]
Dedicated CCO	0.247	0.348**	0.420***	-0.083	0.057	-0.126
	[1.53]	[2.54]	[3.41]	[0.59]	[0.58]	[1.08]
Majority Emp. Owned	-0.089	-0.110	0.026	0.220**	0.252**	-0.022
	[0.88]	[1.14]	[0.28]	[2.26]	[2.23]	[0.17]
Log (Avg. Acct. Size)	-0.100***	-0.089***	-0.076***	-0.056**	-0.034	-0.061
	[3.90]	[4.18]	[3.29]	[2.14]	[1.02]	[1.61]
Percent Client Agents	0.004***	0.003**	0.004***	0.002*	0.003*	0.001
	[3.03]	[2.57]	[2.98]	[1.74]	[1.72]	[0.50]
Hedge Fund Clients	0.006	0.098	0.150	0.127	-0.091	-0.070
	[0.02]	[0.46]	[0.79]	[0.61]	[0.44]	[0.28]
Log (AUM)	0.091***	0.080***	0.062***	0.041*	0.028	0.046
	[4.24]	[4.49]	[3.21]	[1.94]	[1.08]	[1.62]
Log (Firm Age)	0.019	0.007	-0.003	0.009	0.017	0.000
	[1.00]	[0.24]	[0.13]	[0.36]	[0.72]	[0.00]
Model chi-square	138.0***	138.7***	135.3***	78.1***	44.6***	63.9***
Observations	7,352	7,747	8,562	9,088	10,862	10,383
Panel B: Within sample predictions						
# Fraud	104	116	115	83	59	40
Fraud predicted	39	45	37	22	11	10
	37.5%	38.8	32.2	26.5	18.6	25.0
Clean firms	7,248	7,631	8,447	9,005	10,803	10,343
Clean firm false positives	362	381	422	450	540	517
	5.0%	5.0	5.0	5.0	5.0	5.0

Table 5

Initiation versus continuance of fraud

The sample consists of 53,994 firm-year observations. The independent variables are taken from each firm's Form ADV filings as of August 1 each year from 2001 through 2006. Panel A shows the results of a sequential logit regression predicting fraud. The first column shows estimates of the probability that a firm initiates a fraud in the subsequent year. The second column shows estimates of the probability that a firm with a preexisting fraud continues that fraud into the subsequent year. In the interest of brevity the constants are not reported. 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 initiated fraud cases that could be predicted within sample.

Panel A: Predicting initiation versus continuance of fraud		
	Initiate	Continue
Past Fraud	0.537	0.200
	[1.31]	[0.33]
Past Affiliated Fraud	-0.417	0.350
	[-1.33]	[0.65]
Past Regulatory	0.707***	0.010
	[3.87]	[0.03]
Past Civil or Criminal	0.447**	-0.177
	[2.14]	[-0.53]
Referral Fees	0.240	-0.839***
	[1.50]	[-2.78]
Interest in Transaction	0.509***	0.153
	[2.59]	[0.53]
Soft Dollars	-0.124	0.389
	[-0.70]	[1.62]
Broker in Firm	0.349**	-0.562**
	[2.05]	[-2.10]
Investment Company Act	0.612***	-0.023
	[3.04]	[-0.07]
Custody	0.246	0.317
	[1.33]	[1.13]
Dedicated CCO	-0.199	0.305
	[-1.14]	[1.00]
Majority Employee Owned	-0.022	-0.141
	[-0.13]	[-0.43]
Log (Avg. Acct. Size)	-0.181***	0.083
	[-4.63]	[1.10]
Percent Client Agents	0.009***	-0.003
	[3.56]	[-0.51]
Hedge Fund Clients	0.034	-0.335
	[0.09]	[-0.53]
Log (AUM)	0.155***	-0.068
	[4.78]	[-1.15]
Log (Firm Age)	0.014	0.092
	[0.34]	[1.50]
Model chi-square	199.7***	18.3

Panel B: Within sample predictions	
# Fraud	87
Fraud predicted	33
	37.9%
# Clean firms	53,907
Clean firm false positives	2,673
	5.0%

Table 6

Firm-wide versus rogue employee fraud

The sample consists of 53,994 firm-year observations. The independent variables are taken from each firm's Form ADV filings as of August 1 each year from 2001 through 2006. Panel A shows the results of a multinomial probit regression predicting fraud. In the first column, the dependent variable equals one for firms that experience a firm-wide fraud in the subsequent year. In the second column, the dependent variable equals one for firms that experience a rogue employee fraud in the subsequent year. The excluded category is clean firms. In the interest of brevity the constants are not reported. 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. The columns in Panel B correspond to the columns in Panel A. Panel B shows the proportion of fraud that could be predicted within sample.

Panel A: Predicting firm-wide versus rogue employee fraud		
	Firm	Rogue
Past Fraud	-0.698	1.107**
	[1.00]	[2.03]
Past Affiliated Fraud	-0.700*	0.105
	[1.72]	[0.18]
Past Regulatory	0.697***	0.888**
	[3.45]	[2.09]
Past Civil or Criminal	0.066	1.600***
	[0.25]	[3.21]
Referral Fees	0.194	0.926
	[1.15]	[1.59]
Interest in Transaction	0.435**	1.185*
	[2.10]	[1.69]
Soft Dollars	-0.03	-0.925*
	[0.16]	[1.88]
Broker in Firm	0.327*	0.88
	[1.85]	[0.94]
Investment Company Act	0.656***	0.351
	[3.03]	[0.74]
Custody	0.340*	-0.729
	[1.73]	[1.49]
Dedicated CCO	-0.312	0.301
	[1.59]	[0.67]
Majority Employee Owned	-0.013	0.026
	[0.07]	[0.05]
Log (Avg. Acct. Size)	-0.144***	-0.294***
	[3.25]	[3.27]
Percent Client Agents	0.009***	0.003
	[3.45]	[0.41]
Hedge Fund Clients	0.001	-0.441
	[0.00]	[0.40]
Log (AUM)	0.121***	0.372**
	[3.27]	[2.45]
Log (Firm Age)	0.005	0.176
	[0.12]	[0.93]
Model chi-square	144.3***	193.8***

Panel B: Within sample predictions		
# Fraud	450	67
Fraud predicted	109	50
	24.2%	74.6%
# Clean firms	53,477	
Clean firm false positives	2,673	
	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 the Form ADV sample with the TASS hedge fund (TASS), CRSP Mutual Fund (CRSP), and PSN institutional fund (PSN) databases. Each investment adviser's fraud risk is defined as the predicted value from the regression reported in column three of Table 3. We assign this measure of fraud risk to each fund managed by that investment adviser. For each of the return databases, we create two equally weighted portfolios based on the funds' predicted fraud risk. The high fraud risk portfolio includes all funds advised by firms whose predicted fraud risk is greater than the 95^{th} percentile of clean firms. The low fraud risk portfolio includes all other funds. We estimate alphas using monthly returns for each portfolio. We use the Fung and Hsieh (2001) model for the TASS sample and the Carhart (1997) model for the CRSP and PSN samples. For the CRSP mutual fund and PSN Informa databases, we include only equity funds. 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 management and incentive fees. The dependent variable for the CRSP 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 fund is advised by a firm whose predicted value from column three of Table 3 is greater than the 95^{th} percentile of clean firms. Log(Fund AUM) is the logarithm of assets under management for the fund. Log(Fund Age) is the logarithm of fund age in years. Fund Offshore equals one if the fund is registered offshore (i.e., non-US). Leverage equals one if the fund uses leverage. Log(Firm AUM) is the logarithm of assets under management for the firm to which the fund belongs. Turnover % is the annual percentage turnover of the fund's portfolio. Index Fund equals one if the fund is an index fund. Style Fixed Effects are created using Primary Category from TASS, Lipper Objective in CRSP, and Investment Style and Market Cap from PSN. Year Dummies and constant included but not reported. The variables are sampled annually 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	0.0046***	0.0034***	-0.0012**
	[5.96]	[4.24]	[2.11]
CRSP	-0.0021***	-0.0021***	-0.00002
	[5.52]	[4.43]	[0.13]
PSN	0.0001	0.0001	0.00005
	[0.14]	[0.20]	[0.19]

Panel B: Fees				
	TASS		CRSP	PSN
	Management	Incentive	Expense Ratio	Management
High Risk	-0.089	0.022	0.047	-0.004
	[1.32]	[0.04]	[1.40]	[0.25]
Log (Fund AUM)	0.008	-0.017	-0.077***	-0.0004
	[0.88]	[0.22]	[10.21]	[0.14]
Log (Fund Age)	-0.055***	-0.221	0.045***	
	[3.72]	[1.56]	[2.76]	
Fund Offshore	0.115***	0.042		
	[3.54]	[0.15]		
Leverage	0.050*	1.818***		
	[1.76]	[5.04]		
Log (Firm AUM)			-0.024***	-0.004
			[3.07]	[1.07]
Turnover %			0.0005***	0.0002**
			[4.02]	[2.31]
Index Fund			-0.740***	
			[11.80]	
Style Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	5,998	5,998	20,767	8,682
Adjusted R^2	0.141	0.428	0.233	0.227

Table 8
Point-in-time tests using publicly accessible data

Each column uses only the contemporaneously accessible cross-section of Form ADV filings as of August 1 of that year. Panel A shows the estimates from cross-sectional probit regressions. The dependent variable equals one for firms which have a publicly observed prior history of fraud (fraud occurred and was detected between January 1, 1996 and August 1 of the year in which the independent variables are observed). The independent variables reflect the publicly accessible data as of August 1 of the year in the column. 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. Panel B shows the proportion of fraud that could be predicted in year T+1, using Form ADV data at time T as inputs to the prediction model in the aligned column of Panel A.

Panel A: Point-in-time cross-sections						
	Aug '01	Aug '02	Aug '03	Aug '04	Aug '05	Aug '06
Past Affiliated Fraud	0.757***	0.774***	0.111	-0.006	-0.067	-0.188
	[2.70]	[3.14]	[0.39]	[0.03]	[0.30]	[0.87]
Past Regulatory	0.705***	0.976***	0.984***	0.820***	0.953***	1.166***
	[3.99]	[4.66]	[6.56]	[5.24]	[7.40]	[8.69]
Past Civil or Criminal	0.642***	0.445**	0.594***	0.620***	0.409**	0.464***
	[2.85]	[2.28]	[3.44]	[3.59]	[2.53]	[3.50]
Referral Fees	0.437**	0.347*	0.333**	0.234	0.137	0.127
	[2.17]	[1.91]	[2.14]	[1.64]	[1.15]	[1.07]
Interest in Transaction	0.366	0.447**	0.019	-0.251	0.063	0.081
	[1.56]	[2.10]	[0.11]	[1.39]	[0.43]	[0.65]
Soft Dollars	-0.401**	-0.311*	-0.210	-0.342**	-0.102	0.017
	[2.14]	[1.91]	[1.32]	[2.26]	[0.85]	[0.14]
Broker in Firm	0.093	0.045	-0.090	-0.209*	-0.157	-0.133
	[0.54]	[0.25]	[0.69]	[1.71]	[1.26]	[1.05]
Investment Co. Act	0.206	0.155	-0.088	0.284*	0.118	0.143
	[0.88]	[0.71]	[0.43]	[1.65]	[0.66]	[0.87]
Custody	-0.092	-0.019	-0.039	0.165	-0.067	-0.007
	[0.48]	[0.11]	[0.23]	[1.11]	[0.55]	[0.06]
Dedicated CCO	-0.321	-0.178	-0.035	0.174	0.089	0.098
	[0.86]	[0.78]	[0.15]	[1.05]	[0.78]	[0.92]
Majority Emp. Owned	-0.114	0.102	0.011	-0.056	0.083	0.148
	[0.51]	[0.55]	[0.06]	[0.38]	[0.67]	[1.16]
Log (Avg. Acct. Size)	-0.161***	-0.114***	-0.044	-0.062*	-0.069**	-0.067**
	[3.52]	[3.09]	[1.09]	[1.78]	[2.10]	[2.13]
Percent Client Agents	0.000	-0.001	-0.001	0.002	0.001	0.000
	[0.12]	[0.22]	[0.26]	[0.86]	[0.28]	[0.06]
Hedge Fund Clients	0.384	0.074	-0.079	-0.122		-0.421
	[0.93]	[0.20]	[0.20]	[0.32]		[1.08]
Log (AUM)	0.121***	0.095***	0.037	0.057*	0.057**	0.048*
	[3.39]	[2.86]	[1.20]	[1.92]	[2.07]	[1.85]
Log (Firm Age)	0.056*	0.064	0.183***	0.160***	0.124**	0.197***
	[1.70]	[1.30]	[3.29]	[2.79]	[2.20]	[3.99]
Model chi-square	224.8***	116.2***	115.1***	150.1***	139.4***	155.7***
Observations	7,352	7,747	8,562	9,088	9,110	10,383
Panel B: Out-of-sample predictions						
# Fraud	104	116	115	83	59	40
Fraud predicted	28	33	27	20	12	10
	26.9%	28.4	23.5	24.1	20.3	25.0
# Clean firms	7,248	7,631	8,447	9,005	10,803	10,343
Clean firm false positives	360	374	419	456	470	552
	5.0%	4.9	5.0	5.1	4.4	5.3

Table 9

Predictions using a panel of all prior years

Each column represents the predictions an investor could have made at a specified point-in-time had historical Form ADV been publicly available. For each column, the sample consists of a panel of all previously available annual Form ADV filings as of August 1 of each year. (E.g., In Aug 2002, the independent variables are taken from the Aug 2001 cross-section of Form ADV filings. In Aug 2003, the independent variables are taken from the Aug 2002 and Aug 2001 samples of Form ADV filings.) 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. 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 fraud that could be predicted in year T+1, using Form ADV data at time T as inputs to the prediction model in the aligned column of Panel A.

Panel A: Panel of all prior years					
	Aug '02	Aug '03	Aug '04	Aug '05	Aug '06
Past Fraud	0.280	0.226	0.102	0.178	0.254
	[0.82]	[0.94]	[0.37]	[0.79]	[1.21]
Past Affiliated Fraud	-0.321	-0.271*	-0.237*	-0.219*	-0.187
	[1.23]	[1.83]	[1.91]	[1.89]	[1.52]
Past Regulatory	0.187*	0.213***	0.261***	0.288***	0.293***
	[1.89]	[3.53]	[3.42]	[3.82]	[4.15]
Past Civil or Criminal	0.239*	0.220***	0.176*	0.211**	0.181*
	[1.73]	[2.61]	[1.80]	[2.25]	[1.96]
Referral Fees	0.041	0.031	0.045	0.061	0.084
	[0.45]	[0.67]	[0.83]	[1.15]	[1.53]
Interest in Transaction	0.265***	0.257***	0.250***	0.227***	0.216***
	[2.78]	[4.92]	[4.27]	[3.42]	[3.29]
Soft Dollars	-0.037	-0.053	-0.041	-0.034	-0.033
	[0.40]	[1.04]	[0.74]	[0.61]	[0.61]
Broker in Firm	0.202**	0.162***	0.127**	0.117*	0.111*
	[2.33]	[2.86]	[2.04]	[1.93]	[1.89]
Investment Co. Act	0.245**	0.289***	0.301***	0.304***	0.276***
	[2.43]	[4.17]	[4.60]	[4.68]	[3.59]
Custody	0.006	0.048	0.094	0.075	0.088
	[0.06]	[0.72]	[1.20]	[1.07]	[1.37]
Dedicated CCO	0.247	0.305***	0.346***	0.199	0.027
	[1.53]	[3.10]	[3.46]	[1.29]	[0.22]
Majority Emp. Owned	-0.089	-0.099*	-0.053	-0.005	0.016
	[0.88]	[1.90]	[0.71]	[0.06]	[0.19]
Log (Avg. Acct. Size)	-0.100***	-0.094***	-0.087***	-0.079***	-0.072***
	[3.90]	[7.75]	[5.85]	[4.68]	[4.01]
Percent Client Agents	0.004***	0.004***	0.004***	0.003***	0.003***
	[3.03]	[4.18]	[4.29]	[3.90]	[3.94]
Hedge Fund Clients	0.006	0.072	0.114	0.107	0.053
	[0.02]	[0.52]	[0.87]	[0.83]	[0.42]
Log (AUM)	0.091***	0.084***	0.076***	0.067***	0.061***
	[4.24]	[7.72]	[5.46]	[4.39]	[3.89]
Log (Firm Age)	0.019	0.015	0.008	0.008	0.007
	[1.00]	[1.44]	[0.68]	[0.71]	[0.69]
Model chi-square	138.0***	543.8***	304.8***	218.7***	205.5***
Observations	7,352	15,099	23,661	32,749	43,611

Panel B: Out-of-sample predictions					
# Fraud	116	115	83	59	40
Fraud predicted	48	38	21	12	10
	41.4%	33.0	25.3	20.3	25.0
# Clean firms	7,631	8,447	9,005	10,803	10,343
Clean firm false positives	386	402	496	603	517
	5.1%	4.8	5.5	5.6	5.0

Fig. 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.

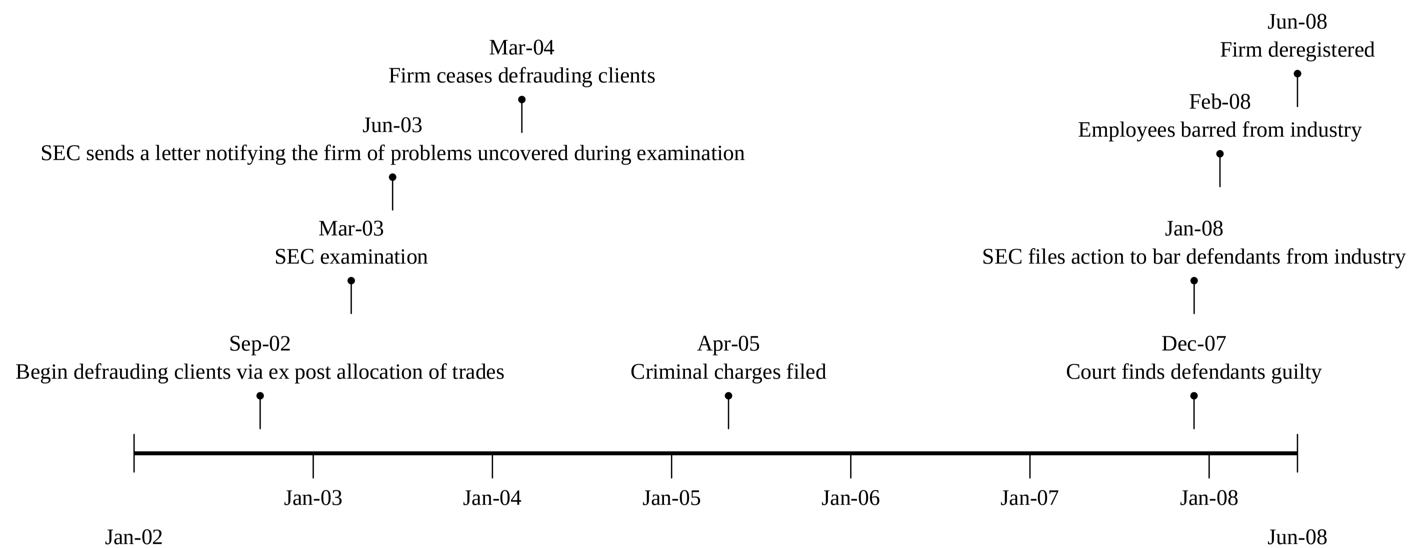
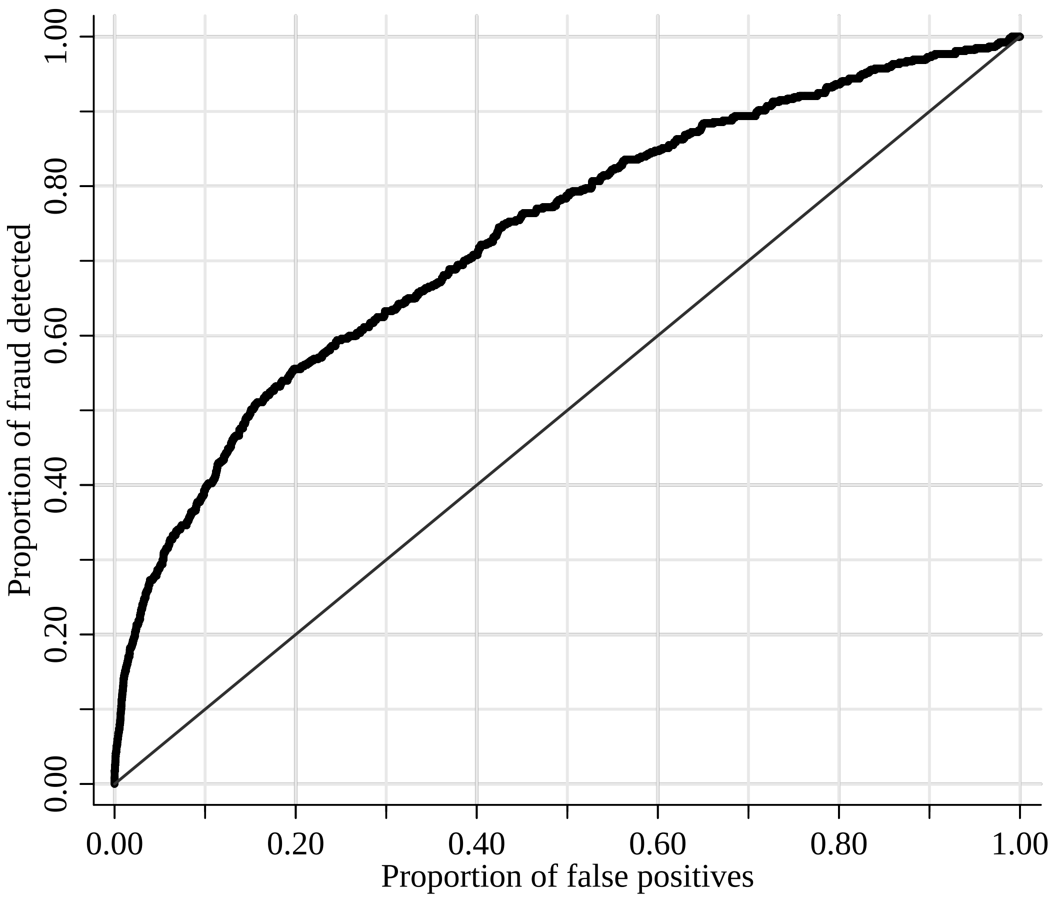


Fig. 2. Proportion of fraud predicted for all false positive rates. This figure shows the receiver operating characteristic (ROC) curve for the probit regression results from the second column of Table 3. The ROC curve shows the relation between the proportion of fraud detected and the proportion of false positives for all possible classification cut-points. The ROC curve is generated by taking each observation's estimated fraud probability, computing the sensitivity and false positives using that point as a cut-point, and then plotting the results.



Appendix
Variable definitions

Variable	Definition	Data source
Past Fraud	The firm committed a previously detected fraud	SEC administrative proceeding or litigation release was filed for firm prior to August 1 of firm-year observation
Past Affiliated Fraud	An affiliate of the firm committed a previously detected fraud	SEC administrative proceeding or litigation release was filed for affiliated firm prior to August 1 of firm-year observation and Schedule D Section 7.A reports fraud firm as affiliate
Past Regulatory	Filed a regulatory disclosure reporting page (DRP)	One of more of: Item 11c1-3, 11d1-5, 11e-4
Past Civil or Criminal	Filed a criminal or civil DRP	One of more of: Item 11a1-2, 11b1-2, 11h1a, 11h1b, 11h1c, 11h2
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: buy (or sell) securities from advisory clients; recommend securities in which you have an ownership interest or serve as underwriter, general or managing partner or have any other sales interest	One of more of: Item 8a1, 8a3, 8b2, 8b3
Soft Dollars	Do you or any related person receive research or benefits other than execution from a broker-dealer or a third party in connection with client securities transactions?	Item 8e
Broker in Firm	Employs registered representatives of a broker-dealer	Item 5b2>0
Investment Company Act	Investment adviser (or sub-adviser) to an investment company registered under the Investment Company Act	Item 2a4
Custody	Do you or any related person have custody of any advisory clients' cash or securities?	One of more of: Item 9a1-2, 9b1-2
Dedicated CCO	CCO has no other stated role within firm	CCO on Schedule A has no other "Title or Status"
Majority Employee Owned	Over 50% aggregate employee ownership	Imputed using Dimmock et al. (2011) method
Log (Avg. Acct. Size)	Logarithm of assets under management per client	Log (Item 5f2c/(Item 5f2f+1)+1)
Percent Client Agents	Percent of banking, mutual, pension, charitable, corporate, and government clients	Sum of items: 5d3, 5d4, 5d5, 5d7, 5d8, 5d9 imputed using Dimmock et al. (2011) method
Hedge Fund Clients	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 (years since date registered with the SEC)

References

Becker, G., 1968. Crime and punishment: An economic approach. The Journal of Political Economy 76, 169–217.

Bollen, N., Pool, V., 2008. Conditional return smoothing in the hedge fund industry. Journal of Financial and Quantitative Analysis 43, 267–298.

Bollen, N., Pool, V., 2009. Do hedge fund managers misreport returns? Evidence from the pooled distribution. Journal of Finance 64, 2257–2288.

Bollen, N., Pool, V., 2010. Predicting hedge fund fraud with performance flags. Unpublished working paper. Vanderbilt University.

Brown, S., Goetzmann, W., Liang, B., Schwarz, C., 2008. Mandatory disclosure and operational risk: Evidence from hedge fund registration. Journal of Finance 63, 2785–2815.

Brown, S., Goetzmann, W., Liang, B., Schwarz, C., 2009. Estimating operational risk for hedge funds: The ω-score. Financial Analysts Journal 65, 43–53.

Busse, J., Goyal, A., Wahal, S., 2010. Performance persistence in institutional investment management. Journal of Finance 65, 765–790.

Carhart, M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57–82.

Cassar, G., Gerakos, J., 2010. Determinants of hedge fund internal controls and fees. Accounting Review 85, 1887–1919.

Darby, M., Karni, E., 1973. Free competition and the optimal amount of fraud. Journal of Law and Economics 16, 67–88.

Dechow, P., Ge, W., Larson, C., Sloan, R., 2011. Predicting material accounting misstatements. Contemporary Accounting Research 28, 17–82.

Dechow, P., Sloan, R., Sweeney, A., 1996. Causes and consequences of earnings manipulation: An analysis of firms subject to enforcement actions by the SEC. Contemporary Accounting Research 13, 1–36.

Dimmock, S., Gerken, W., Marietta-Westberg, J., 2011. What determines the allocation of managerial ownership within firms? Unpublished working paper. Nanyang Technological University.

Fung, W., Hsieh, D., 2001. The risk in hedge fund strategies: Theory and evidence from trend followers. Review of Financial Studies 14, 313–341.

Karpoff, J., Lott, J., 1993. The reputational penalty firms bear from committing criminal fraud. Journal of Law and Economics 36, 757–802.

Karpoff, J., Lott, J., Wehrly, E., 2005. The reputational penalties for environmental violations: Empirical evidence. Journal of Law and Economics, 68, 653–675.

Khorana, A., Servaes, H., Tufano, P., 2009. Mutual fund fees around the world. Review of Financial Studies 22, 1279–1310.

Klein, B., Leffler, K., 1981. The role of market forces in assuring contractual performance. Journal of Political Economy 89, 615–641.

Lott, J., 1996. The level of optimal fines to prevent fraud when reputations exist and penalty clauses are unenforceable. Managerial and Decision Economics 17, 363–380.

Murphy, D., Shrieves, R., Tibbs, S., 2009. Understanding the penalties associated with corporate misconduct: An empirical examination of earnings and risk. Journal of Financial and Quantitative Anaylsis 44, 55–83.

Newey, W., West, K., 1987. A simple, positive semi-definite, hetroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Straumann, D., 2009. Measuring the quality of hedge fund data. Journal of Alternative Investments 12, 26–40.

Zitzewitz, E., 2006. How widespread was late trading in mutual funds? American Economic Review 96, 284–289.