Do Investors Have Valuable Information About Brokers?[1]

Hammad Qureshi† Jonathan Sokobin†

August 2015

Abstract

We examine the value of information available to investors through BrokerCheck: the most comprehensive source of information about brokers' professional background and regulatory history that helps investors make informed choices about which brokers to use. We do so by assessing the predictability of investor harm associated with brokers based on BrokerCheck information. We find that BrokerCheck information, including disciplinary records, financial disclosures, and employment history of brokers has significant power to predict investor harm. The 20% of brokers with the highest *ex-ante* predicted probability of investor harm are associated with more than 55% of the investor harm cases and the total dollar investor harm in our sample. Our findings suggest that investors have access to valuable information that allows them to discriminate between brokers with a high propensity for investor harm from other brokers. We also assess the impact of releasing additional non-public information on BrokerCheck and find that investors may benefit from information about harm associated with brokers' coworkers.

Keywords: BrokerCheck, Disclosures, Investor harm, CRD

JEL Classification: G2, G19, G20, G28, G29, K20, K22

[1] The views expressed in this paper are those of the authors and do not necessarily reflect the views of FINRA or of the authors' colleagues on FINRA staff. We are grateful to Chester Spatt for conducting a peer-review of the paper and providing valuable comments. We would also like to thank Viral Acharya, Ozzy Akay, Michael Goldstein, Charles Jones, Pete Kyle, Tian Liang, Jonathan Macey, Gideon Saar and seminar participants of the 2015 FINRA Economic Advisory Committee Meeting for their comments. We are grateful to FINRA staff for invaluable insights into the organization and history of the CRD data and outstanding technology support.

†Office of the Chief Economist, FINRA, 1735 K Street NW, Washington, DC 20006. Email: ChiefEconomist@finra.org.

1. Introduction

The brokerage industry in the United States represents one of the largest segments of the U.S. financial services sector.[2] At the end of 2014, the revenue generated by the brokerage firms exceeded $200 billion dollars.[3] Brokerage firms have more than 160,000 branch offices that employ more than 630,000 individual brokers. These brokers offer financial advice to and transact a variety of securities on behalf of millions of investor households.

To help investors make informed choices about the brokers with whom they conduct business, the Financial Industry Regulatory Authority (FINRA) provides an online tool, BrokerCheck, to investors. BrokerCheck provides information on the professional background, including disciplinary history and customer complaints, of more than 1.2 million current and former brokers.[4] FINRA describes BrokerCheck as an important tool for enhancing investor protection and encourages investors to use it just as consumers readily use online tools, such as Yelp or Trip Advisor to compare service providers in other industries.[5] More than 29 million broker searches were conducted on BrokerCheck in 2014, with approximately 18.9 million summary records viewed and approximately 7 million downloads of detailed reports on brokers.[6] BrokerCheck represents the single most complete source of information about brokers available to the public.[7]

The information FINRA makes available through BrokerCheck is derived from its Central Registration Depository (CRD®), a central licensing and registration system for the U.S. securities industry. The CRD system contains qualification, employment and disciplinary records of brokers and firms and

---

[2] In this paper, brokers refer to individual representatives who are registered with FINRA, and brokerage firms or firms refer to FINRA registered broker-dealer firms.

[3] Based on information reported by FINRA members on their Financial and Operational Combined Uniform Single (FOCUS) filings.

[4] A description of BrokerCheck can be found on FINRA's website at: http://brokercheck.finra.org. BrokerCheck provides users access to information about individual brokers and brokerage firms. This paper focuses on the information content related to individual brokers only. We use the term brokers and registered representatives (RR) interchangeably in this paper.

[5] *See*, *e.g*., remarks by Richard G. Ketchum, Chairman and Chief Executive Officer of FINRA, delivered to the Consumer Federation of America Consumer Assembly, March 14, 2013. The remarks can be found at http://www.finra.org/newsroom/speeches/031413-remarks-consumer-federation-america-consumer-assembly. An important difference between these types of tools, which are primarily crowd-sourced reviews by consumers, and BrokerCheck is that the information on BrokerCheck comes from required filings with securities regulators, and made by brokerage firms and individual brokers rather than from investors. FINRA rules prescribe the content, format and timing of information that must be disclosed.

[6] Based on BrokerCheck usage statistics compiled by FINRA staff as of year-end 2014. BrokerCheck is not only used by investors but also by firms and industry professionals. For example, brokerage firms also use BrokerCheck to screen candidates as part of the recruiting process.

[7] Certain states also make publicly available information about brokers licensed to do business in their state. However, state regulators differ on what information is released because each state is governed by its own public records laws, which differ from state to state. In addition, most states only provide information about brokers licensed by that state.

FINRA makes a significant portion of this information available to the public through BrokerCheck.[8] The type and amount of CRD information FINRA releases to the public, is governed by its BrokerCheck Disclosure Rule and instructions from the SEC. FINRA has revised this rule several times in the last decade to expand the scope of information available on BrokerCheck.[9] Nonetheless, BrokerCheck does not include certain CRD information about brokers, such as some financial events and performance on qualification examinations.

Given that BrokerCheck is considered to be the most comprehensive source of information available to investors about brokers' professional histories, it is important to examine the value of BrokerCheck information to investors and to assess whether BrokerCheck would be enhanced by the inclusion of additional non-public information.[10] This paper is in part motivated by public comments that have questioned the value of information available to investors through BrokerCheck.[11]

In this paper, we examine the following research questions: Do investors have access to valuable information about brokers through BrokerCheck today? Would expanding the information provided by BrokerCheck to include other non-public information required to be filed in CRD enhance the value of BrokerCheck to investors?

To address these questions, we construct an annual panel of information from 2000 to 2013 about brokers who likely have direct dealings with the public. The panel includes 181,133 such brokers who registered with FINRA in 2000 or later and tracks their information since their first registration. The panel includes data publicly released on BrokerCheck as well as other non-public CRD data. To our knowledge, the data used in this paper represents the most comprehensive dataset on brokers used in an academic study, and allows us to contribute to the economically important but not well-studied literature on the brokerage industry.

To assess the value of information available to investors through BrokerCheck, we examine the predictability of investor harm associated with brokers based on BrokerCheck information. We measure investor harm using complaints filed by customers against their brokers and their subsequent outcomes. Since some customer complaints may lack merit or suitable evidence of investor harm, we only count complaints that led to awards against brokers or settled above a *de minimis* threshold. This allows us to focus our analysis on outcomes that are likely associated with material investor harm. Less than 1.5% of the brokers in our sample meet this definition of being

---

[8] *See* "Study and Recommendations on Improved Investor Access to Registration Information about Investment Advisers and Broker-Dealers", January 2011 (SEC Study) for a description of CRD and information available on BrokerCheck.
[9] *See* SEC Study, 17-19.
[10] For example, would BrokerCheck be more informative to investors if it were to include information on bankruptcies that are more than 10 years old and satisfied judgments and liens? Would qualification exam scores and the number of times brokers failed those exams enhance the information content available to investors through BrokerCheck?
[11] *See, e.g.*, "PIABA Warning: Finra withholds critical "red flag" information in broker background check disclosures," March 6, 2014, and "Stockbrokers Who Fail Test Have Checkered Records," Wall Street Journal, April 14, 2014. These public commenters claim that certain information about brokers not disclosed on BrokerCheck is indicative of investor harm and should be made available to investors.

associated with investor harm in the fourteen-year panel. In this context, harm does not imply malfeasance on the part of the broker. Instead it only suggests that a third party (regulator, arbitrator or the firm) considered the claim to be worthy of remuneration.

To evaluate the impact of including additional sets of non-public information on BrokerCheck, we test the incremental power of such information to predict investor harm above and beyond the "baseline" of what is currently on BrokerCheck. The four sets of non-public information we evaluate relative to the "baseline" are: (1) investor harm associated with other brokers at firms where the broker is registered (*i.e.,* harm associated with coworkers or "HAC"), to proxy the compliance culture at these firms, (2) currently undisclosed financial events, including satisfied liens and bankruptcies more than 10 years old, (3) undisclosed disciplinary events, including internal reviews, and closed or dismissed regulatory actions, investigations and civil judicial actions, and (4) performance on qualification exams, including exam scores and proportion of exams failed.

We find that the information currently available to investors through BrokerCheck, including disciplinary records, financial and other disclosures, and employment history, has significant power to discriminate between brokers associated with investor harm events and other brokers. The 20% of brokers with the highest *ex-ante* predicted probability of investor harm are associated with more than 55% of the investor harm events in our sample. The proportion of total dollar harm represented by these harm events is more than 55.5 percent suggesting that our predictions capture economically meaningful events and not merely small cases. We also examine the trade-off between investor harm events predicted correctly (true positives) and harm events predicted incorrectly (false positives). Our out-of-sample tests and sensitivity analyses to alternative measures of investor harm confirm the robustness of our predictions. We stress, however, that prediction does not imply a causal relation between the disclosed information and investor harm. Overall, our results suggest that BrokerCheck provides valuable information to investors, thereby allowing them to discriminate between brokers with a high propensity for investor harm from other brokers.

With respect to the impact of releasing additional non-public CRD information on BrokerCheck, we find that HAC leads to an economically meaningful increase in the overall power to predict investor harm, in the context of our model. Undisclosed financial events, undisclosed disciplinary events or exam performance, however, do not enhance the overall predictability of investor harm. These results suggest that investors would benefit from information on harm associated with brokers' coworkers.

Our findings are subject to certain limitations. First, although we find that certain broker characteristics can predict investor harm, we cannot infer that these characteristics cause harm. Prediction does not imply causality, as broker characteristics may be jointly determined with the decision to harm investors. In other words, these broker characteristics may be endogenous. However, because our goal is prediction rather than establishing causality, the potential endogeneity of these broker characteristics does not change our interpretation. Second, as with any prediction model, only detected investor harm events can be included in the analysis. Although we conduct several out-of-sample predictions and sensitivity tests for alternative harm measures, and these tests confirm that our predictions are robust, we cannot rule out the possibility that the predictions may be biased because undetected investor harm events are unobservable. Third, although we approximate and include a subset of likely "public-facing" brokers based on the number of state registrations, we cannot rule out the possibility that our predictions may be biased because our sample excludes other public-facing brokers, or includes certain non-public facing brokers, with different characteristics. Finally, our use of prediction models is not intended to suggest that BrokerCheck is envisioned to be used for predicting investor harm. Instead, we use predictive models only as a tool to evaluate the value of information currently available to investors on BrokerCheck and other information collected in CRD.

The rest of the paper is organized as follows. Section 2 discusses the related literature. Section 3 describes the data and our research approach. In Section 4, we assess whether investors have access to valuable information about brokers through BrokerCheck. In Section 5, we evaluate the impact of including additional sets of non-public CRD information on BrokerCheck. Section 6 provides our conclusion.

2. Related research

Predicting performance or propensity for misconduct by individuals has been the subject of research across various academic fields. For example, studies in medicine use information on physician characteristics to predict medical malpractice claims. Gibbons et al. (1994) find that a physician's age, gender, specialty, prior claims, and risk management education are important predictors of malpractice claims. Tamblyn et al. (2007) find that a physician's scores on national clinical skills examinations are significant predictors of complaints to medical regulatory authorities. Similarly, literature on criminal recidivism uses information on prisoner characteristics to predict the likelihood of their return to prison.[12]

In the finance literature, a few papers have developed methods to detect or predict investor harm by investment advisory firms.[13] Bollen and Pool (2010) examine hedge funds' manipulation of reported returns and find that suspicious return patterns can predict fraud charges. Dimmock and Gerken (2012) test the predictability of investment fraud based on mandatory disclosures in the Form ADV

[12] *See*, *e.g*., Schmidt and Witte (1987).

[13] Papers in the accounting and corporate finance literature examine financial misconduct associated with corporations. Karpoff, Koester, Lee and Martin (2011) provide a literature review on these papers. These papers focus on understanding the causes and consequences of financial misconduct by corporations (*e.g*., the impact of financial misrepresentation or accounting restatements by corporations on their stock prices). Some papers also develop methods to predict financial misconduct, such as accounting misstatements by corporations (*e.g.,* Dechow, Larson and Sloan (2007), and Price, Sharp and Wood (2011)). These papers differ from our study, in part, because they examine misconduct associated with corporations as opposed to individuals.

filings by investment managers. The authors find that disclosures related to past regulatory and legal violations, conflicts of interest, and monitoring have significant power to predict fraud. Brown, Goetzmann, Liang, and Schwarz (2009) examine the value of Form ADV disclosures in assessing the operational risk of hedge funds. The authors test whether operational risk can predict hedge fund closure, flows and returns. Overall, their findings suggest that hedge funds operated by managers who filed Form ADV had better past performance and had more assets than those operated by managers who did not file. The authors also find a strong positive association between potential conflicts identified in the Form ADV filing and past legal and regulatory problems.

This line of finance research focuses on developing methods to test the predictability of harm associated with investment management *firms*, as a whole, as opposed to *individual* investment managers or other financial professionals, which is the focus of our study. To our knowledge there are no papers in this literature that examine investor harm associated with individual brokers and test the relevance or significance of certain information about these brokers and their propensity for harm.[14]

3. Data and methodology

This study uses data collected in FINRA's CRD. CRD is the securities industry registration and licensing database that was implemented by FINRA in 1981 in order to consolidate a multi-state, paper-based registration process into a single, nationwide filing system. In 1999, FINRA introduced "Web CRD," which allowed electronic filing of registration forms through its website. Information in CRD is obtained through the Uniform Forms that brokers, brokerage firms and regulators complete as part of the securities industry registration and licensing process.[15]

The Uniform Forms in CRD contain information about qualification, employment and disciplinary records of brokers and firms. CRD information is generally self-reported by the brokerage firms and

---

[14] Some papers in the Computer Science literature have applied machine learning algorithms to detect and predict frauds by individuals using explicit social-network data (*e.g.*, Fawcett and Provost (1997), Cortes et al. (2001), and Hill et al. (2006)). A paper in this literature that is related to our study, Neville et al. (2005), provides an application of relational learning algorithms (a sub-discipline in artificial intelligence and machine learning) to predict securities fraud by brokers. The authors find that networks of relationships between brokers can help in identifying securities fraud and that their model predictions are highly correlated with the subjective evaluations of experienced NASD examiners. While Neville et al. (2005) also examine investor harm associated with brokers, the focus of their study is to use relational knowledge discovery models to rank brokers based on the propensity of harm. Our focus, on the other hand, is to test the value of certain information about brokers in predicting investor harm, using econometric methods that are well-established in the finance and economics literature.

[15] Six different Uniform Forms are used to file information with CRD: (1) Form U4 (Uniform Application for Securities Industry Registration or Transfer); (2) Form U5 (Uniform Termination Notice for Securities Industry Registration); (3) Form U6 (Uniform Disciplinary Action Reporting Form); (4) Form BD (Uniform Application for Broker-Dealer Registration), an SEC form; (5) Form BDW (Uniform Request for Broker-Dealer Withdrawal), also an SEC form; and (6) Form BR (Uniform Branch Office Registration Form).
See http://www.finra.org/industry/web-crd/current-uniform-registration-forms-electronic-filing-web-crd for information on the Uniform Forms.

brokers[16] but incorrect or missing reports can trigger regulatory action by FINRA.[17] FINRA rules require brokers and brokerage firms to keep their registration data accurate and up-to-date by updating CRD no later than 30 days after they learn that an update is required and in some instances, within 10 days.[18]

We use a subset of CRD data during the 2000-2013 period. Specifically, our sample includes all brokers who first registered in 2000 or thereafter, the year after Web CRD was introduced in mid-1999.[19] We end our sample in 2013 to allow sufficient time for most customer complaints to reach a resolution, such as a settlement or an award.[20] Focusing on this sample allows us to track information, including employment and disciplinary histories since the first registration for each broker. CRD includes information on all registered representatives, including public-facing brokers as well other brokers that generally do not deal with public investors.[21] Currently, CRD forms do not collect information about the role a broker plays within a firm that could be used to distinguish public-facing brokers from other brokers. In order to approximate and exclude brokers that do not generally provide services to public investors, we restrict our sample to brokers who held more than three state registrations for at least half of their registration tenure.[22] Our sample includes 181,133 brokers who registered with FINRA in 2000 or later, and likely have direct dealings with the public.

To construct an annual panel for these brokers for the predictive regressions, we aggregate disclosure events and other information for each broker during each calendar year in the 2000-2013

---

[16] Regulators also provide information to CRD, such as information on qualification exams or information on certain disciplinary actions.

[17] FINRA rules require firms to investigate the business reputation, qualifications and experience of job applicants before the firms apply to register these job applicants with FINRA. These rules also require firms to have taken appropriate steps to verify the accuracy and completeness of the information contained in the Uniform Forms before they are filed. The SEC recently adopted a FINRA-proposed rule that requires firms to adopt written procedures that are designed to verify the accuracy and completeness of the information contained in an applicant's Form U4 before it is filed. (*See FINRA Regulatory Notice 15-05* at http://www.finra.org/industry/notices/15-05.) As part of this rule proposal, FINRA has been conducting background searches of financial public records on all registered persons and searches of criminal public records on a risk-based basis on any registered person who has not been fingerprinted within the past five years. Nonetheless, CRD data used in this paper may still contain errors and omissions, which could affect the interpretation of our results.

[18] *See* Article V, Section 2(c) of the FINRA By-Laws.

[19] As discussed above, CRD data goes back to the 1980s or earlier. However, prior to 1999 the data was stored in a legacy system, which was based on paper registration. While the legacy system was partially converted to Web CRD in 1999, we use the post-1999 data to avoid any time inconsistencies in information arising from system conversions.

[20] As discussed in more detail below, we measure the occurrence of investor harm based on customer complaints that resulted in a non-*de minimis* settlement or an award. Although most of these complaints are resolved within a year, some may span more than a year. To allow sufficient time for customer complaints to reach a resolution, we end our sample in 2013.

[21] These non-public facing registered representatives include proprietary traders, product wholesalers, as well as compliance, operations and support staff. As noted above, we use the term registered representatives and brokers interchangeably throughout the paper.

[22] CRD includes information on state registrations by brokers. Based on its experience, FINRA staff believes that brokers with more than three state registrations generally deal with the public investors.

period. Disclosure events in CRD are often associated with multiple filings.[23] To avoid double-counting disclosures due to multiple filing sources, we use disclosure occurrence data compiled by FINRA disclosure review staff that review and aggregate disclosure events across forms into "unique" occurrences. Many disclosure events in CRD are also associated with multiple dates that span several years and involve multiple actions. For such disclosures, we use the earliest date when the underlying event was reported.[24]

FINRA releases certain CRD information about brokers to the public through BrokerCheck. BrokerCheck includes information on broker qualifications, employment history and disclosure events. Certain disclosures in CRD are included on BrokerCheck when they are reported but subsequently removed after a specified period of time or after a certain resolution. For example, BrokerCheck includes bankruptcy disclosures[25] for the first 10 years and excludes them after they are more than 10 years old.[26] In order to evaluate the value of information included on BrokerCheck and the impact of including additional information to it, we need to separate disclosure events that are disclosed on BrokerCheck at any point in time from those that are not. We do so by constructing historical "at the time views" of BrokerCheck during the 2000-2013 period. Specifically, for each disclosure event we calculate when it was included on BrokerCheck and if and when it was excluded, based on the dates and resolution of the underlying event. For example, we split bankruptcy disclosures into: i) bankruptcies less than 10 years old, and ii) bankruptcies more than 10 years old, and for each year in our annual panel we check whether a particular bankruptcy event was more or less than 10 years old in that year, and count it accordingly.

3.1. Measures of investor harm and broker characteristics

i. Investor harm

We measure the occurrence of investor harm based on complaints customers filed against the broker that result in a non-*de minimis* settlement or an award to an investor.[27] Brokers are required to

[23] For example, a customer complaint is reported by the broker on Form U4 and if the broker was subsequently terminated, the same complaint could also be reported by the firm in Form U5.

[24] As noted above, the FINRA By-Laws require brokers and registered representatives to report any disclosure event within 30 days after they learn about it and in some instances within 10 days. These rules ensure that there is not a significant lag in when the underlying event occurred and when it is reported. There are sometimes inconsistencies in dates reported across forms (*e.g.*, in U4 and U5) for the same underlying disclosure event. In such cases, FINRA staff selects the dates from what it considers as the most reliable source. We apply the same logic in selecting dates across forms.

[25] The term bankruptcy as used in this paper refers to bankruptcies, Securities Investor Protection Corporation (SIPC) events, and compromises with creditors.

[26] Similarly, BrokerCheck includes information in CRD on judgments and liens when they are not satisfied and excludes them after they are satisfied.

[27] An alternative measure of investor harm could be based on regulatory actions. However, there would be certain limitations with such a measure. First, CRD only contains information on the date when a regulatory action was initiated, which could be several years after the actions associated with investor harm occurred or were detected. These lags

submit information on all customer complaints to CRD for inclusion on BrokerCheck. Complaints may be resolved through settlements or awards, or may lead to enforcement or other legal actions, or they may remain unresolved. These complaints may remain unresolved because they may lack merit or suitable evidence.[28] In order to focus the analysis on outcomes that are likely associated with material customer harm, we only count the complaints that led to an award against the broker or settled above a *de minimis* threshold. We first use the CRD settlement threshold for reporting customer complaints on Uniform Forms of $10,000 for complaints that settled prior to May 18, 2009 and $15,000 for settlements thereafter. To account for the possibility that some firms may still treat settlement at these dollar levels as *de minimis*, we also consider an alternative measure based on a higher $25,000 threshold.[29]

Customer complaints that lead to an award or settlement are associated with multiple CRD filings that may span multiple years.[30] For example, a complaint filed by a customer in 2005 may lead to arbitration and subsequently result in an award in 2006.[31] CRD tracks information about individual complaints as they evolve over time but does not contain information on the timing of the underlying occurrence of investor harm (*e.g.,* the start or duration of actions associated with investor harm). We proxy the occurrence of investor harm based on complaint filing year.[32] For instance, in the above example we define investor harm as occurring in 2005, and use the prior year's (2004) information to predict it.[33] By predicting the occurrence of complaint in 2005, rather than its resolution in 2006, we

between the occurrence of investor harm and the filing of regulatory action can bias the predictive regressions. Second, regulatory actions may only capture a small proportion of investor harm events. Finally, regulatory actions may include actions against brokers that are not associated with any direct harm to the investor. We note that while we do not measure investor harm based on regulatory actions, information about regulatory actions is already disclosed on BrokerCheck and included as an explanatory variable in predicting investor harm, amongst other predictors, as discussed in more detail below.

[28] For example, customers may file complaints that are false or erroneous. Erroneous complaint filings are generally subsequently withdrawn by customers, dismissed by firms or may be closed by the arbitration panel in favor of the broker. Brokers may also choose to settle erroneous complaints, for a *de minimis* amount, to avoid litigation costs. FINRA staff estimates that for the complaints filed in the 2000-2014 period approximately 52% of the (non-pending) complaints were dismissed by firms, withdrawn by customers or closed by the arbitration panels.

[29] This threshold corresponds to approximately the 25th percentile of settlement amount in our sample. We also considered higher thresholds or $50,000 and $100,000. Our results do not change materially with these higher thresholds.

[30] Most of the customer complaints lead to an award or settlement within a year of being filed. For example, 75% of the complaints that led to an award or settled above the CRD threshold reached a resolution within a year. Approximately 20% of the complaints resolved in the second year, whereas the resolution of the remaining 5% took more than 2 years.

[31] CRD aggregates all the events associated with a complaint into a "single" complaint occurrence and tracks the evolution of the complaint over time.

[32] For complaints that are initiated as litigations or arbitrations, we use the earliest available date for such litigation or arbitration (*e.g.,* arbitration notice date).

[33] While CRD does not contain information on the underlying actions associated with investor harm, FINRA Rule 4530 requires firms to report quarterly summary information (including the start and end date of the underlying actions) regarding written customer complaints. CRD customer complaints and 4530 complaints essentially capture the same set of complaints but because they are governed by different rules, the reporting requirements differ and, as a result, it is not possible to map the two sets of complaints. However, we can use 4530 complaints to approximate the extent of lag between the occurrence of investor harm and customer complaint received date for complaints in our sample. We find

avoid potential biases caused by a correlation between resolution and time variation in the predictive variables.

Table 1 summarizes the distribution of investor harm events during the 2000-2013 period. We report the number of brokers associated with 0, 1 or more investor harm events during their tenure, for different measures of investor harm. These measures are based on customer complaints that led to an award and complaints that settled above the specified threshold.

Table 1 shows that a large majority of brokers (over 98.5%) in our sample are not associated with investor harm. Most of the brokers associated with investor harm only had one complaint that led to an award or settled above a *de minimis* threshold during the period. The distribution of investor harm is similar for both the CRD threshold and the $25,000 threshold.

ii. Broker characteristics

To evaluate the value of information disclosed on BrokerCheck and the impact of including additional information on it, we construct separate measures for broker characteristics that are disclosed on BrokerCheck and those that are not. Table 2 presents a summary of broker characteristics that are disclosed on BrokerCheck in Panel A, and those that are not disclosed in Panel B. The table summarizes characteristics for the last year for which we have information for each broker (and thereby incorporates the most recent information in the sample for each broker).[34] The first column presents averages for all brokers, whereas the second and third columns present means for brokers associated with any event leading to investor harm and brokers without any such association, respectively. The last two columns report the difference in means between broker with and without association with investor harm and the (univariate) statistical significance of this difference, respectively.

Judgments and Liens (Unsatisfied) correspond to the number of judgments and liens against the broker that were not satisfied until the latest year of the broker's tenure in our sample. Bankruptcy (<10 years) captures the number of bankruptcies filed by the broker within the last 10 years of the year under consideration. Disclosed disciplinary events are the sum of regulatory actions, investigations, civil judicial actions, and terminations that are included on BrokerCheck. BrokerCheck also includes criminal-related disclosures that are reportable on the current Uniform Forms. Exams passed captures the number of qualification exams (Series 6, 7, 63 and 66) a broker passed.[35] The

---

that the duration of most investor harm-related actions is less than a month and the lag between the initiation of such actions and the complaint received date is less than a year.

[34] This corresponds to 2013 information for brokers that were registered in 2013, and the last year of registration for brokers that de-registered prior to 2013. For example if a broker was registered from 2000 until 2010, the table includes 2010 information for that broker.

[35] We include exams that are generally associated with brokers who deal with public investors. Series 6 and Series 7 are FINRA registration exams and Series 63 and Series 66 are state exams for brokers registering with a state to conduct business with the public. A description of these examinations can be found at: http://www.finra.org/industry/qualification-exams .

next four characteristics are based on broker employment information that is disclosed on BrokerCheck. Association with an expelled firm is an indicator that equals 1 for brokers who were registered at a firm that has been expelled by FINRA for disciplinary or other reasons.[36] Brokers need not have been associated at the time the firm was expelled for the indicator to be set to 1; association at any time prior to the expulsion is sufficient. Number of prior employers corresponds to the number of firms that the broker has worked at and left until the latest year of his/her tenure. Dual Registration is an indicator that equals 1 for brokers who are also registered with the SEC as investment advisors. Gender is an indicator that equals 1 for male brokers and 0 for female brokers.

Panel A in Table 2 shows that brokers associated with investor harm have a higher average number of unsatisfied liens, bankruptcies (within the last 10 years), disciplinary and criminal events, and greater number of exams passed and prior employers than those who are not associated with investor harm, and these differences are statistically significant at the 1% level. In addition, a greater proportion of brokers associated with investor harm are males, have dual registration, and are associated with expelled firms. Information about these events and broker characteristics is available to investors through BrokerCheck.

Panel B of Table 2 summarizes information on four sets of broker characteristics that are not currently disclosed on BrokerCheck. These include undisclosed financial events, undisclosed disciplinary events, information about exam performance and our measure of compliance culture, HAC. Undisclosed financial events include judgments and liens that have been satisfied and bankruptcies that are more than 10 years old, respectively.[37] Undisclosed disciplinary events account for regulatory actions, investigations and civil judicial actions, against the brokers that were closed without action, or dismissed by the year under consideration. It also includes internal reviews, regardless of their resolution. The next two characteristics are based on brokers' performance on Series 6, 7, 63 and 66 qualification exams. Exams failed capture the number of times a broker failed the qualification exams and exam scores represent the cumulative average score across the four exam series. We compute HAC for each broker in a given year by calculating the average number of

[36] Firms may be expelled by FINRA as a result of disciplinary actions or for other reasons, such as failure to pay registration fees. BrokerCheck does not distinguish disciplinary expulsions from other expulsions. To be consistent with the information that is released on BrokerCheck, we count all expulsions in our measure for association with expelled firms.

[37] We note that at the time of initial registration, FINRA rules require new brokers to only provide information about bankruptcies that have occurred within the last 10 years. Thus, all brokers do not have equivalent information in CRD. For example, consider two individuals who ultimately become brokers and who both declared bankruptcy in 1995. If broker A initially registered with FINRA in 2000, the 1995 bankruptcy would have been reported. Our bankruptcy measure would then count the event as a bankruptcy of under 10 years until 2004 and a bankruptcy of over 10 years thereafter. If broker B first registered with FINRA in 2008, the 1995 bankruptcy would not have been reported to FINRA and would not be reflected in broker B's history. As a result, our predictions are based on partial CRD information on bankruptcies that are more than 10 years old. Similarly, at the time of initial registration, brokers are only required to provide information on judgment and liens that are unsatisfied. Accordingly, our predictions for satisfied liens are also based on partial CRD information. These data limitations in CRD, however, do not affect the interpretation of our predictions because our goal is to evaluate the value of information that is already collected in CRD.

investor harm events per registered representative (RR) for all other RRs at the same firm, averaged over the last five years across all firms with which the broker is employed in the year under consideration.[38]

Panel B in Table 2 shows that brokers associated with investor harm have a higher average number of satisfied liens, bankruptcies (more than 10 years old), undisclosed disciplinary events, number of exams failed, and HAC score than those not associated with investor harm. These differences are statistically significant at the 1% level. In addition, brokers associated with investor harm have on average lower cumulative exam scores than other brokers.

Table 2 shows that there are statistically significant differences in characteristics between broker associated with investor harm and those without such association, suggesting that these broker characteristics may be *individually* informative (*e.g.,* in a univariate test) for discriminating between brokers. The univariate statistical tests in Table 2, however, do not capture the association of these characteristics with the propensity of investor harm while controlling for other characteristics. Neither do these tests capture the ability of these characteristics to predict investor harm, individually or collectively, or the trade-off between predicting correctly and incorrectly. We discuss our methodology to test the predictability of investor harm based on these characteristics in the next sub-section.

3.2. Methodology

To assess whether investors have access to valuable information, through BrokerCheck, that allows them to evaluate a broker's propensity for investor harm, we test the predictability of harm based on BrokerCheck information. We predict the probability of investor harm for a given broker with particular characteristics by estimating the following probit model:

$$P(Y_{it} = 1|X_{it-1}) = G(\alpha_0 + \beta_1 X_{it-1}^{BrokerCheck}) \quad [1]$$

where "P" denotes the probability of investor harm in year *t* and *G* represents the assumed normal probability distribution. We estimate the model using prior year's information on BrokerCheck, $X_{it-1}^{BrokerCheck}$, to predict the next year's probability of investor harm, $P(Y_{it} = 1|X_{it-1})$, [39] for each broker *i* (i=1,2,…N) for each year *t* (t=1,2,….,n_i). We estimate pooled probit models and report

[38] Certain states allow brokers to be registered with multiple firms at the same time, whereas other states either restrict registrations to a single firm or to affiliated firms. In our sample, less than 2% of the brokers are registered with multiple (affiliated or unaffiliated) firms at the same time. For these brokers, we give equal weight to investor harm events per RR for all firms. Similarly, when brokers move across firms within a year, we assign equal weights to all firms with which they are registered in that year.

[39] An indicator for investor harm; Y_{it} equals 1 (for broker 'i') in the year one or more investor harm events are reported, and 0 otherwise.

standard errors clustered by broker in the paper.[40] The model also includes prior year's market return. While market returns are not displayed on BrokerCheck, public investors readily have access to this information, and so we include it as a control in our prediction models.

To evaluate the predictive ability of the BrokerCheck information, we calculate an Investor Harm Score (IH-Score) for each broker-year observation, by dividing its predicted probability based on regressions in equation [1] by the unconditional probability of investor harm.[41] An IH-Score of 1.00 indicates that the broker has the same likelihood of being associated with investor harm as the overall sample. IH-Scores less (greater) than one indicate lower (higher) probabilities of investor harm. This terminology closely follows Dechow, Ge, Larson, and Sloan (2007), who create an ''F-score'' to detect accounting fraud. We then sort and rank each broker-year observation into quintiles based on the IH-Score and evaluate the frequency with which brokers associated with investor harm fall into high quintiles.

We also analyze the receiver operating characteristic (ROC) curve based on the predictive regressions. As discussed in more detail below, ROC curves are generated non-parametrically by taking each broker-year observation's predicted value from the probit model as a cut-point, and then computing both the proportion of investor harm events correctly predicted as well as the false positives. The ROC curve accounts for the trade-off between investor harm events predicted correctly (true positives) and harm events predicted incorrectly (false positives) by the model. False positives occur when the model incorrectly predicts that a broker not associated with investor harm will be associated with investor harm in the subsequent year.

To evaluate the impact of including additional sets of non-public CRD information on BrokerCheck, we include additional information in the "baseline" predictive regression in [1] and test the incremental predictive power of such information above and beyond the baseline. The model can be formally represented by the following equation:

$$P(Y_{it} = 1 | X_{it-1}) = G(\alpha_0 + \beta_1 X_{it-1}^{BrokerCheck} + \beta_2 X_{it-1}^{NOT\ in\ BrokerCheck}) \quad [2]$$

[40] We also considered panel models that account for unobserved broker effects (*i.e.,* broker characteristics that cannot be directly observed but affect the likelihood of investor harm). Specifically, we accounted for unobserved effects using probit models with random-effect. Our results based on random effect probit models do not change materially. We note that random-effect probit models assume that the unobservable effects are not correlated with broker characteristics. Generally, one way to account for possible correlation between observable and unobservable broker characteristics is to use a fixed effect model. However, fixed effects probit models cannot be estimated consistently, because of the "incidental parameters" problem (*i.e.*, β_1 in equation [1] cannot be consistently estimated with fixed T and $N \to \infty$ in the fixed effects framework). For details, *see* Wooldridge, "Econometric Analysis of Cross Section and Panel Data", Chapter 15.

[41] The unconditional probability equals to the number of broker-year observations associated with investor harm divided by the total number of observations in our sample.

where $X_{it-1}^{NOT\ in\ BrokerCheck}$ includes the four sets of non-public CRD information about brokers (HAC, undisclosed financial events, undisclosed disciplinary events, and exam performance). We test the incremental predictive power of these additional sets of information, by including them individually or in combination with other sets of information to the baseline model in [1].

4. Do investors have access to valuable information about brokers through BrokerCheck?

In this section, we assess whether BrokerCheck provides valuable information to investors by testing the predictability of investor harm based on BrokerCheck information. The purpose of these tests is prediction in order to assess the information content and, as noted previously, we make no claims regarding causality. Broker characteristics that are predictive of investor harm may be endogenous but because our goal is prediction rather than establishing causality, the potential endogeneity of the predictors does not influence our interpretation.

A caveat in interpreting our findings is that we observe only detected investor harm events.[42] We address the issue of undetected investor harm by conducting extensive out-of-sample tests to ensure the predictions, and hence the economic content of the information, are robust.

4.1. Prediction models

Table 3 shows the results of panel probit regressions that predict investor harm based upon BrokerCheck and other publicly available information ("baseline" predictions).[43] The sample is an unbalanced panel of 1,014,873 broker-year observations during the 2000-2013 period. As discussed above, we measure the occurrence of investor harm based on customer complaints that led to an award or settled above a threshold. Table 3 reports results for both the settlement thresholds we consider—the CRD threshold and the $25,000 threshold are shown in column 1 and 2, respectively. The dependent variable equals 1 if one or more investor harm events occur in the next calendar year.[44]

[42] Observed investor harm depends on factors, such as the unobservable true rate of harm, the probability of detection given a fixed level of monitoring, and the allocation of monitoring resources. Ideally, the probit model will predict the true rate of investor harm. However, if certain predictive variables are correlated with either monitoring or detection, this relation could affect the interpretation of the results.

[43] As discussed above, the baseline predictions include prior year's market return, as a control. While market returns are not displayed on BrokerCheck, public investors readily have access to this information.

[44] The Z-scores shown in brackets are based on standard errors clustered by broker. The chi-square tests at the bottom of each column show the significance of the overall model.

Past awards and settlements correspond to the number of customer complaints against the broker that led to an award or settlement above the specified threshold.[45] To prevent a "look ahead" bias in the results arising from the time-interval between complaint filing and its resolutions, this measure only includes complaints that were known to have resulted in an award or settled (above specified threshold) by the end of the prior year. For example, for a complaint that was filed in 2005 and resulted in an award against the broker or settled above a *de minimis* amount in 2006, we only count such a complaint, starting in 2006. Past awards and settlements have a statistically significant positive coefficient, suggesting that brokers with past complaints that resulted in award or settlement have a higher propensity for investor harm in the subsequent year. Past financial events, including unsatisfied judgments and liens and bankruptcies within the last 10 years have a positive association with investor harm. Past disclosed disciplinary events and past criminal events are also positively associated with investor harm.

Number of qualification exams passed is not associated with investor harm. The next four characteristics are based on broker employment history that is disclosed on BrokerCheck. Past association with expelled firms, number of prior employers, employment years, and past dual registration are all positively associated with investor harm. This suggests that brokers associated with expelled firms, dual-registered brokers, and those with greater number of prior employers and more employment years have a higher propensity for investor harm in the subsequent year.[46]

Gender is also associated with investor harm. All else equal, male brokers have a higher propensity for investor harm. Prior market return captures the annual return of the market (S&P 500 index). While market returns are not displayed on BrokerCheck, public investors readily have access to this information, and so we include it as a control in our baseline predictions. Market return is intended to control for the effect of market conditions on the propensity for customers to file complaints. For example, customers may be more likely to file complaints and arbitrations against brokers if the markets are down and the performance of their investment portfolio is adversely affected. Complaints in these scenarios are often related to suitability of the product for the customer and whether the broker provided appropriate information about the risks associated with the investment. A positive coefficient on this variable would imply not only that there is a greater

[45] Our measure is based on the total number of relevant customer complaints as opposed to an indicator that only accounts for the existence of prior complaint(s). Accounting for the number of complaints allows us to capture the marginal impact of each additional complaint on the probability of investor harm. The marginal impact is assumed to be constant over the number of complaints (*i.e.,* the effect of a previous investor harm increases or decreases the probability of investor harm by the same amount), which we consider to be a reasonable approximation of the effect of previous complaints.

[46] This finding is consistent with a prior SEC study that reviewed the hiring, retention, and supervisory practices of nine of the largest U.S. broker-dealers. The study found that brokers associated with investor harm were able to move between firms freely after customers registered complaints, suggesting a positive association between number of prior employers and investor harm. (*See* The Large Firm Project, A Review of Hiring, Retention and Supervisory Practices, Divisions of Market Regulation and Enforcement, United States Securities and Exchange Commission, May 1994, available at http://www.sec.gov/news/studies/rogue.txt.)

propensity for customer complaints after market losses, but also that these complaints are also more likely to resolve with a non-*de minimis* award or settlement. This hypothesis is supported by results in Table 3, which show that prior market return has a negative association with investor harm.

4.2. Predictive power of models

The predictive regressions in Table 3 indicate that BrokerCheck information has a statistically significant relation with subsequent investor harm. This finding is important, but the key question of interest in our study is whether investors have access to valuable information that allows them to evaluate broker information and discriminate between brokers with potentially higher propensity for investor harm from other brokers. To address this question, we sort and rank each broker-year observation into quintiles based on its scaled probability (IH-Score), discussed above, and evaluate the frequency with which brokers associated with investor harm and brokers without such association fall into each quintile. If BrokerCheck information has no predictive ability then brokers associated with investor harm and other brokers will be randomly dispersed across quintiles, implying that 20% of both sets of brokers would be allocated to each IH-Score quintile. In contrast, if BrokerCheck information is useful in discriminating between brokers with a high propensity of investor harm from other brokers, we expect the highest proportion of brokers associated with investor harm to be clustered in the highest quintile and the lowest proportion of these brokers in the lowest quintile.

Panel B in table 3 shows that, based on CRD threshold in Column 1, 55.5% of the investor harm events are in Quintile 5 compared to only 3.8% in Quintile 1. Further, the percentages of investor harm events increase monotonically from Quintile 1 to Quintile 5. The results are similar for the alternative $25,000 settlement threshold, as shown in Column 2. Table 3 also reports the proportion of total dollar harm that falls into each quintile. We winsorize the dollar harm at the 99^{th} percentile to prevent the reported results being driven by large outliers. The total dollar harm, based on aggregating all awards and settlements above the *de minimis* threshold in our sample, is over $500 million, in nominal dollars. The proportion of total dollar harm that fall into Quintile 5, based on CRD and $25,000 settlement thresholds is 55.7% and 55.8%, respectively. This indicates that our predictions capture economically meaningful investor harm cases and not merely small cases.

We also examine the tradeoff between correctly predicted outcomes of investor harm events associated with individual brokers (true positives) and cases where the model incorrectly predicts investor harm (false positives), by analyzing the receiver operating characteristic (ROC) curve based on the predictive regressions in equation [1]. False positives can be interpreted as the opportunity cost to investors of erroneously limiting the pool of brokers with whom they may conduct business. Although failing to predict investor harm would be more costly for investors than mistakenly avoiding a broker not associated with investor harm (false positive cases), an investor would need to avoid multiple false positive cases for every true positive case avoided.[47]

To illustrate the possible tradeoffs between false positives and predicted investor harm (true positives), Figure 1 shows a receiver operating characteristic (ROC) curve for the baseline prediction model in the first column 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 investor harm events correctly predicted as well as the false positives. Random prediction of true and false investor harm would result in a straight 45-degree line.

The ROC curve captures the full range of all possible tradeoffs between the prediction of investor harm and false positives. Following Dechow, Ge, Larson & Sloan (2011) and Dimmock & William (2012), we provide greater detail for one possible tradeoff; the proportion of investor harm events that could be predicted (true positive rate) and the corresponding false positive rate at IH-Score cutoff of 1. As discussed above, IH-Score of 1.00 indicates that the broker has the same likelihood of being associated with investor harm as the overall sample. This cutoff implies that all brokers with predicted probability greater than unconditional probability of a randomly picked broker in the sample (which, as a rule of thumb, can be considered to be "above normal risk" brokers) are predicted to be associated with harm. For example, as shown in Figure 1, at IH-Score cutoff of 1.00, 71.1% of the investor harm events can be predicted correctly at a false positive rate of 33.0%.

In Panel B of Table 3, we report additional statistics about true positives and false positives at this IH-Score cutoff. Panel B shows that the baseline model (with CRD settlement threshold) predicts 1,889 of 2,656 investor harm events (71.1%) at a false positive rate of 33.0% (the model incorrectly predicts investor harm in 334,126 out of 1,012,217 non-investor harm events). The corresponding proportion of total dollar harm predicted based on the baseline model is higher at 73.5% (compared to 71.2%), confirming that these predictions capture economically meaningful investor harm cases.

Overall, these results suggest that the information currently available to investors through BrokerCheck, including disciplinary records, financial and other disclosures, and employment history, has significant power to predict investor harm. The 20% of brokers with the highest *ex-ante* predicted probability of investor harm are associated with more than 55% of the investor harm events and the total dollar investor harm in our sample. Next, we assess the robustness of these results to out-of-sample validation tests.

[47] For example, as shown in Table 3 and discussed below, to avoid 1,889 true positive cases, an investor would need to avoid 334,126 false positive cases. In other words, an investor would need to avoid approximately 176 false positive cases for every true positive.

4.3. Out of sample validation tests

In this subsection, we test whether the predictions in Table 3, are robust out-of-sample. We do so by performing K-fold cross-validation tests over the 2000-2013 period. Each model is estimated on a randomly selected subsample of brokers, and the coefficient estimates from this subsample are used to classify brokers in the hold-out sample. Specifically, each broker in the sample is randomly assigned to one of 10 groups.[48] 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. We repeat this process 20 times, for a total of 200 hold-out samples.

The results for the out-of-sample validation tests, shown in Table 4, indicate that the predictive power of the models is only slightly lower in the hold-out samples. For example, the baseline predictions (based on CRD settlement threshold) allocate 55.5% of the investor harm events within-sample to Quintile 5 (Table 3) compared to an average of 55.3% of investor harm events in the hold-out samples (Table 4). Similarly, the baseline predictions at IH-Score cutoff of 1.00 correctly predicted 71.1% of investor harm events within-sample, compared to an average of 70.9% investor harm events in the hold-out samples. The K-fold test predicts a minimum of 70.7% and a maximum of 71.3% investor harm events across the 20 repetitions. These results suggest that the model is quite stable. Overall, results of the K- fold cross validation tests support the robustness of our baseline predictions in Table 3.

In sum, we find that the information currently available to investors through BrokerCheck has significant power to predict investor harm, within the context of the models tested. Our out-of-sample tests and sensitivities to alternative measures of investor harm events confirm the robustness of these predictions. Overall, these results suggest that BrokerCheck provides valuable information to investors that permits them to discriminate between brokers who are likely to be associated with investor loss events and other brokers.

5. Impact of including additional information on BrokerCheck

In this section we evaluate the impact of including additional sets of non-public information about brokers, already collected within CRD, on BrokerCheck. We do so by including additional information in the baseline predictive regression and testing the incremental power of such information to predict investor harm above and beyond the baseline. The four sets of non-public CRD information we evaluate are our measure of investor harm associated with coworkers; undisclosed financial events, including satisfied liens and bankruptcies more than 10 years old; undisclosed disciplinary

[48] We randomly assign brokers, and not broker-years, to avoid overstating the results due to non-independence.

history, including internal reviews, and closed or dismissed regulatory actions, investigations and civil judicial actions; and broker exam performance, including exam scores and proportion of exams failed.

Table 5 shows the impact of including additional information to the baseline prediction. The first column reproduces the "baseline" prediction (based on CRD threshold) in Table 3, to facilitate comparison. The next 4 columns show the effect of including additional sets of information, one at a time.

HAC has a statistically significant positive coefficient, implying that there is a positive association between past HAC score and propensity for investor harm. More importantly, including HAC increases the predictive power of the baseline model, as shown in the Panel B of Table 5. For example, the percentage of investor harm events in the highest quintile increases from 55.5% to 58.9% and the proportion of dollar harm predicted increases from 55.7% to 57.1% when HAC is included in the model. Table 5 shows that including HAC also increases the predictive power at the IH-Score cutoff of 1. The proportion of true positives increases from 71.1% to 74.0% while that of false positives decreases from 33.0% to 31.9%. The corresponding proportion of dollar harm predicted increases from 73.5% to 77.2%. This 3.7% increase in dollar harm predicted corresponds to combined awards or settlements of more than $18 million, which suggests that the increase in predictive power from HAC is also economically important. Overall, these results show that including information about HAC on BrokerCheck would increase the predictability of investor harm.

In the next column, we add information on past undisclosed financial events. Satisfied judgments and liens events and bankruptcies that are more than 10 years old have positive coefficients but these coefficients are not statistically significant at the conventional 5% level. Additionally, including undisclosed financial events does not increase the predictive power of the baseline model. For example, the percentage of investor harm events and the proportion of total dollar harm in the highest quintile or at IH-Score cutoff of 1 stay at essentially the same levels as the baseline model.

Undisclosed disciplinary events are positively associated with investor harm, but including these events does not enhance the overall predictive power of the baseline model. Including undisclosed disciplinary events leads to a slight increase in the proportions of investor harm events and dollar harm in the highest quintile but this increase is not economically significant. In addition, including undisclosed disciplinary events reduces the predictive power at IH-Score cutoff of 1.

Brokers' average exam scores are negatively associated with investor harm but there is no statistically significant association between the number of times a broker failed the exams and investor harm. More importantly, including exam performance generally leads to a reduction in predictive power.

The results in Table 5 show the impact of including the four sets of non-public CRD information individually. To account for potential correlations and interactions within these information sets, and across other BrokerCheck information, we report the impact of including all possible combinations of the non-public CRD information sets in the next table.

Table 6 shows the predictive power of the models, based on investor harm predicted at IH-Score equal to unity, when combinations of additional sets of information are included to the baseline. The first 7 specifications show the impact of including combinations of undisclosed financial events, undisclosed disciplinary events and exam performance. These results show that the additional sets of information reduce the predictive power at the IH-Score cutoff of unity.

The next 8 specifications (specifications 8-15) include HAC with other sets of information. These results confirm that HAC leads to a significant increase in predictive power, both individually and in combination with other information. Similarly, the proportions of investor harm events and dollar harm predicted at IH-Score of unity range from 73.9% to 74.4%, and 76.9% to 77.2%, respectively, compared to the baseline proportions of 71.1% and 73.5%. Overall, these results confirm that information about HAC enhances the overall predictive power to discriminate between brokers associated with investor harm events and other brokers, whereas information on undisclosed financial events and disciplinary events or information on broker's exam performance do not.

We consider the HAC measure as an indicator of "compliance culture" in the sense that it identifies firms that employ a greater number of brokers who were associated with investor harm in the recent past, without distinguishing where that prior investor harm occurred. We recognize that there may be other reasons for a higher HAC, such as a particular firm choosing to sell a specific product that might be shown ex post to have been poorly designed and thus be the subject of many related complaints. Nonetheless, unconditionally our HAC measure should discriminate between firms whose brokers are associated with many investor harm events from other firms. Considering that HAC is measured as the proportion of coworkers associated with harm events, it accounts for differences in the size of the employing firm. We note, however, that the marginal impact of an additional investor harm event will have a larger impact on HAC for smaller firms than for larger firms.

We also compare the incremental predictive power of information about harm associated with coworkers to that of disclosures currently made public on BrokerCheck. We do so, in Table 7, by assessing the impact of replacing disclosed events by HAC (*i.e.,* excluding sets of disclosed events, one at a time, from the baseline prediction and including HAC instead). Table 7 shows that replacing any of the individual financial or disciplinary disclosures provided by BrokerCheck today with our measure of investor harm associated with coworkers leads to an increase in the power of the model to predict investor harm. Overall, these results show that information about harm associated with coworkers is not only important relative to the non-public CRD information we evaluate in this paper, but also significant compared to disclosures that are already released on BrokerCheck.

6. Conclusion

In this study we assess whether BrokerCheck currently provides useful information to investors about brokers. We also evaluate the impact of releasing additional non-public information already collected by FINRA on BrokerCheck. To assess whether BrokerCheck currently provides useful information to investors that allows them to evaluate a broker's propensity for investor harm, we test the predictability of investor harm based on BrokerCheck information. Subsequently, we evaluate the impact of including additional sets of non-public information about brokers on BrokerCheck by testing the incremental predictive power of such information above and beyond what is disclosed on BrokerCheck.

We find that the information currently available to investors through BrokerCheck has significant power to discriminate between brokers associated with investor harm events and other brokers. The 20% of brokers with the highest *ex-ante* predicted probability of investor harm are associated with more than 55% of investor harm events and the total dollar harm in our sample. We stress that prediction does not imply a causal relation between the disclosed information and investor harm, as broker characteristics may be jointly determined with the decision to harm investors. In other words, these broker characteristics may be endogenous. However, because our goal is prediction rather than establishing causality, the potential endogeneity of these broker characteristics does not change our interpretation. Our out-of-sample tests and sensitivities to alternative measures of investor harm confirm the robustness of these predictions. These results suggest that investors have access to valuable information through BrokerCheck that allows them to discriminate between brokers with whom they may conduct business.

With respect to the impact of releasing additional non-public CRD information on BrokerCheck, we find that HAC leads to an economically meaningful increase in the overall predictive power. Undisclosed financial events, undisclosed disciplinary events or exam performance, however, do not enhance the predictability of investor harm. These results suggest that investors may benefit from information about harm associated with brokers' coworkers.

References

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

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.

Cortes, C., Pregibon, D., Volinsky, C., 2001. Communities of interest. Proceedings of the Fourth International Conference on Advances in Intelligent Data Analysis (IDA), 105–114.

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., 2012. Predicting Fraud by Investment Managers. Journal of Financial Economics 105, 153-173.

Fawcett, T., Provost, F., 1997. Adaptive fraud detection. Data Mining and Knowledge Discovery 3, 291–316.

Fawcett, T., Provost, F., 1999. Activity monitoring: Noticing interesting changes in behavior. Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and DataMining, 53–62.

Gibbons, R., Hedeker, D., Charles, S., Frisch, P., 1994. A random-effects probit model for predicting medical malpractice claims. Journal of the American Statistical Association 89(427), 760-767.

Hill, S., Agarwal, D., Bell, R., Volinsky, C., 2006. Building an effective representation for dynamic networks. Journal of Computational & Graphical Statistics 15(3), 584–608.

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

Karpoff, M., Koester, A., Lee, D., Martin, G., 2012. A critical analysis of databases used in financial misconduct research. Unpublished working paper, University of Washington.

Neville, J., Simsek, O., Jensen, D., Komoroske, J., Palmer, K., Goldberg, H., 2005. Using relational knowledge discovery to prevent securities fraud. Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 449–458.

Neville, J., Jensen, D., 2007. Relational dependency networks, Journal of Machine Learning Research 8, 653-692.

Price, R., Sharp, N., Wood, D., 2011. Detecting and predicting accounting irregularities: A comparison of commercial and academic risk measures. Accounting Horizons 25(4), 755-780.

Schmidt, P., Witte, A., 1987. Predicting criminal recidivism using "split population" survival time models. NBER Working Paper # 2445.

Tamblyn, R., Abrahamowicz, M., Dauphinee, D., Wenghofer, E., Jacques, A., Klass, D., ... Hanley, J., 2007. Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. American Medial Asociation 298, 993–1001.

Wooldridge, J., 2002. Econometric analysis of cross section and panel data. MIT Press, Cambridge, MA.

Zitzewitz, E., 2012. Forensic economics. Journal of Economic Literature 50(3), 731-769.

Figure 1: Investor Harm Predicted for All False Positive Rates

The ROC curve shows the relation between the proportion of investor harm events detected and the proportion of false positives for all possible false positive rates.

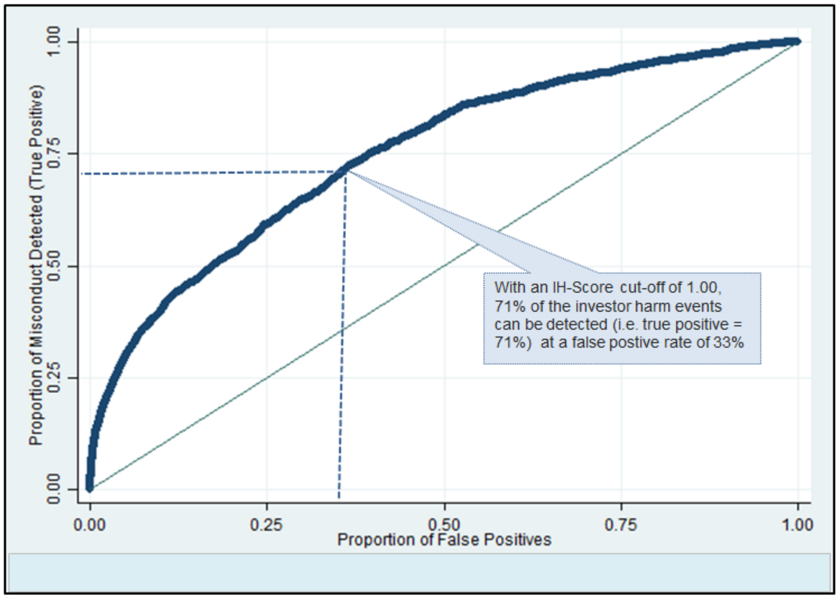

Table 1: Summary of Investor Harm Measures

This table summarizes the distribution by brokers of investor harm measures over the 2000-2013 period. Occurrence of investor harm is based on customer complaints that led to an award and customer complaints that settled above the specified threshold. CRD threshold is $10,000 for complaints that settled before 2009 and $15,000 afterwards.

# of Investor Harm Events	# of Brokers Associated With Investor Harm Based on:	
	Awards & Settlements (CRD threshold)	Awards & Settlements ($25,000 threshold)
0	178,784	179,350
1	1,922	1,457
2	288	217
3 or more	139	109
Total	181,133	181,133

Table 2: Summary of Broker Characteristics

This table summarizes information about brokers' most recent characteristics during the 2000-2013 period. There are 181,133 unique brokers in the sample. Panel A shows the broker characteristics that are disclosed on BrokerCheck, while Panel B shows broker characteristics that are not disclosed on BrokerCheck. For a description of these characteristics, see Appendix A. Column [1] presents average for all brokers, whereas [2] and [3] present averages for brokers associated with investor harm and those without such association, respectively. [4] reports the difference [2] and [3]. [5] reports the (univariate) statistical significance of the difference in [4]. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels.

	[1]	[2]	[3]	[4]=[2]-[3]	[5]
Characteristics	All Brokers	Brokers Associated with Investor Harm	Brokers Not Associated with Investor Harm	Difference	p-value
Panel A: Characteristics Disclosed on BrokerCheck					
Past Settlements and Awards	0.012	0.924	0.000	0.924	0.000***
Judgments and Liens (Unsatisfied)	0.017	0.105	0.016	0.089	0.000***
Bankruptcy Disclosures (< 10 years)	0.054	0.080	0.054	0.026	0.016**
Disclosed Disciplinary Events	0.009	0.095	0.007	0.087	0.000***
Criminal Events	0.026	0.063	0.026	0.038	0.000***
Exams Passed	2.128	2.218	2.127	0.090	0.000***
Association with Expelled Firm	0.004	0.039	0.003	0.036	0.000***
Number of Prior Employers	0.841	1.976	0.826	1.105	0.000***
Employment Years	6.495	10.038	6.449	3.590	0.000***
Dual Registration	0.426	0.663	0.423	0.240	0.000***
Gender (Male)	0.738	0.885	0.736	0.149	0.000***
Panel B: Characteristics NOT Disclosed on BrokerCheck					
Undisclosed Financial Events					
Judgments and Liens (Satisfied)	0.016	0.065	0.015	0.049	0.000***
Bankruptcy Disclosures (> 10 years)	0.013	0.032	0.013	0.019	0.000***
Undisclosed Disciplinary Events	0.005	0.061	0.005	0.056	0.000***
Exam Performance					
Exams Failed	0.465	0.542	0.464	0.078	0.000***
Average Exam Score	79.104	78.447	79.112	(0.665)	0.000***
HAC	0.004	0.010	0.004	0.005	0.000***

Table 3: "Baseline" Predictions Based on BrokerCheck Information

	(1)	(2)
	Awards & Settlements (CRD threshold)	Awards & Settlements ($25,000 threshold)
Panel A: Investor Harm Predictors		
Past Awards & Settlements	0.332***	0.332***
	(8.936)	(7.834)
Past Judgments and Liens (Unsatisfied)	0.112***	0.102***
	(5.748)	(5.012)
Past Bankruptcy Disclosures (< 10 years old)	0.0226**	0.0265***
	(2.324)	(2.643)
Past Disclosed Disciplinary Events	0.230***	0.247***
	(5.038)	(5.954)
Past Criminal Events	0.170***	0.176***
	(6.240)	(5.680)
Exams Passed	0.00663	0.00512
	(0.507)	(0.350)
Past Affiliation with Disciplined Firm	0.432***	0.436***
	(6.779)	(6.436)
Number of Prior Employers	0.0861***	0.0855***
	(15.91)	(14.91)
Employment Years	0.0222***	0.0247***
	(16.59)	(17.12)
Past Dual Registration	0.279***	0.273***
	(18.81)	(16.36)
Gender (Male)	0.292***	0.301***
	(14.28)	(12.83)
Past Market Return	-0.514***	-0.554***
	(-14.84)	(-14.23)
Model chi-square	2303.9***	2019.5***
Observations	1,014,873	1,014,873

Table 3 *continued*

	(1) Awards & Settlements (CRD threshold)	(2) Awards & Settlements ($25,000 threshold)
Panel B: Predictive Power		
IH-Score quintiles and goodness of fit percentages		
IH-Score Lowest Quintile		
No Investor Harm Events	20.0%	20.0%
Investor Harm Events	3.8%	3.3%
Dollar Investor Harm Predicted	4.1%	3.7%
IH-Score Quintile 2		
No Investor Harm Events	20.0%	20.0%
Investor Harm Events	7.9%	7.6%
Dollar Investor Harm Predicted	7.9%	8.0%
IH-Score Quintile 3		
No Investor Harm Events	20.0%	20.0%
Investor Harm Events	11.4%	11.0%
Dollar Investor Harm Predicted	8.4%	9.2%
IH-Score Quintile 4		
No Investor Harm Events	20.0%	20.0%
Investor Harm Events	21.3%	21.5%
Dollar Investor Harm Predicted	23.9%	23.2%
IH-Score Highest Quintile		
No Investor Harm Events	19.9%	19.9%
Investor Harm Events	55.5%	56.7%
Dollar Investor Harm Predicted	55.7%	55.8%
Investor Harm Predicted at IH-Score cutoff =1		
True Positives		
# Investor Harm Events	2,656	1,996
Investor Harm Events Predicted	1,889	1,433
True Positive (%)	71.1%	71.8%
False Positives		
# No Investor Harm	1,012,217	1,012,877
No Investor Harm False Positives	334,126	325,604
False Positive (%)	33.0%	32.1%
Proportion of $ Harm Predicted	73.5%	72.4%

Table 4: Out of Sample Predictions

	Awards & Settlements (CRD threshold)	Awards & Settlements ($25,000 threshold)
Average IH-Score quintiles and goodness of fit percentages		
IH-Score Lowest Quintile		
No Investor Harm Events	20.0%	20.0%
Investor Harm Events	3.8%	3.2%
Dollar Investor Harm Predicted	4.1%	3.7%
IH-Score Quintile 2		
No Investor Harm Events	20.0%	20.0%
Investor Harm Events	8.0%	7.5%
Dollar Investor Harm Predicted	8.0%	8.2%
IH-Score Quintile 3		
No Investor Harm Events	20.0%	20.0%
Investor Harm Events	11.4%	11.0%
Dollar Investor Harm Predicted	8.4%	8.9%
IH-Score Quintile 4		
No Investor Harm Events	20.0%	20.0%
Investor Harm Events	21.5%	21.5%
Dollar Investor Harm Predicted	23.9%	23.1%
IH-Score Highest Quintile		
No Investor Harm Events	19.9%	19.9%
Investor Harm Events	55.3%	56.7%
Dollar Investor Harm Predicted	55.5%	56.0%
Investor Harm Predicted at IH-Score cutoff =1		
True Positives (%)		
Average	70.9%	71.6%
Median	70.9%	71.6%
Min	70.7%	71.4%
Max	71.3%	72.0%
False Positives (%)		
Average	32.9%	32.1%
Median	32.9%	32.1%
Min	32.8%	32.0%
Max	32.9%	32.2%
Proportion of $ Harm Predicted		
Average	73.1%	72.4%
Median	73.0%	72.3%
Min	72.3%	71.7%
Max	74.6%	73.1%

Table 5: Predictions Based on BrokerCheck and Additional Sets of Information

	(1)	(2)	(3)	(4)	(5)
	Baseline	Baseline + HAC	Baseline + Undisclosed Financial Events	Baseline + Undisclosed Disciplinary Events	Baseline + Exam Performance
Panel A: Investor Harm Predictors					
Disclosed on BrokerCheck					
Past Awards & Settlements	0.332***	0.312***	0.332***	0.328***	0.332***
	(8.936)	(9.171)	(8.944)	(9.038)	(9.016)
Past Judgments and Liens (Unsatisfied)	0.112***	0.0951***	0.108***	0.110***	0.110***
	(5.748)	(5.239)	(5.629)	(5.564)	(5.685)
Past Bankruptcy Disclosures (< 10 years old)	0.0226**	0.0241**	0.0222**	0.0228**	0.0223**
	(2.324)	(2.506)	(2.274)	(2.347)	(2.291)
Past Disclosed Disciplinary Events	0.230***	0.198***	0.229***	0.132***	0.229***
	(5.038)	(4.092)	(5.013)	(2.585)	(5.044)
Past Criminal Events	0.170***	0.158***	0.169***	0.173***	0.168***
	(6.240)	(5.772)	(6.216)	(6.357)	(6.145)
Exams Passed	0.00665	0.0217*	0.00660	0.00696	0.00673
	(0.509)	(1.649)	(0.505)	(0.532)	(0.515)
Past Affiliation with Disciplined Firm	0.432***	0.277***	0.432***	0.405***	0.425***
	(6.778)	(4.352)	(6.779)	(6.248)	(6.678)
Number of Prior Employers	0.0861***	0.0782***	0.0860***	0.0837***	0.0843***
	(15.91)	(14.54)	(15.88)	(15.31)	(15.52)
Employment Years	0.0222***	0.0233***	0.0222***	0.0222***	0.0223***
	(16.59)	(17.57)	(16.53)	(16.60)	(16.56)
Past Dual Registration	0.279***	0.277***	0.279***	0.280***	0.281***
	(18.81)	(18.59)	(18.79)	(18.85)	(19.07)
Gender (Male)	0.292***	0.285***	0.292***	0.291***	0.299***
	(14.29)	(13.88)	(14.29)	(14.30)	(14.42)
Not Disclosed on BrokerCheck					
Past HAC		11.91***			
		(23.64)			
Past Judgments and Liens (Satisfied)			0.0218*		
			(1.778)		
Past Bankruptcy Disclosures (> 10 years)			0.0232*		
			(1.954)		
Past Undisclosed Disciplinary Events				0.290***	
				(5.013)	
Exams Failed					-0.00553
					(-0.605)
Past Average Exam Score					-0.0041***
					(-2.776)
Past Market Return	-0.515***	-0.529***	-0.515***	-0.514***	-0.514***
	(-14.84)	(-15.13)	(-14.84)	(-14.83)	(-14.80)
Model chi-square	2303.8***	2892.2***	2309.1***	2383.6***	2327.5***
Observations	1,014,873	1,014,873	1,014,873	1,014,873	1,014,873

Table 5 *continued*

	(1)	(2)	(3)	(4)	(5)
	Baseline	Baseline + HAC	Baseline + Undisclosed Financial Events	Baseline + Undisclosed Disciplinary Events	Baseline + Exam Performance
Panel B: Predictive Power					
IH-Score quintiles and goodness of fit percentages					
IH-Score Lowest Quintile					
Investor Harm Events	3.8%	2.7%	3.8%	3.7%	3.5%
No Investor Harm Events	20.0%	20.0%	20.0%	20.0%	20.0%
Dollar Investor Harm Predicted	4.1%	3.8%	4.1%	4.1%	4.1%
IH-Score Quintile 2					
Investor Harm Events	7.9%	6.8%	7.7%	8.1%	8.1%
No Investor Harm Events	20.0%	20.0%	20.0%	20.0%	20.0%
Dollar Investor Harm Predicted	7.9%	7.0%	7.9%	8.0%	8.2%
IH-Score Quintile 3					
Investor Harm Events	11.4%	10.7%	11.6%	11.4%	11.7%
No Investor Harm Events	20.0%	20.0%	20.0%	20.0%	20.0%
Dollar Investor Harm Predicted	8.4%	7.7%	8.4%	8.3%	8.3%
IH-Score Quintile 4					
Investor Harm Events	21.3%	20.9%	21.3%	21.5%	21.7%
No Investor Harm Events	20.0%	20.0%	20.0%	20.0%	20.0%
Dollar Investor Harm Predicted	23.9%	24.4%	23.9%	23.9%	24.7%
IH-Score Highest Quintile					
Investor Harm Events	55.5%	58.9%	55.6%	55.4%	55.1%
No Investor Harm Events	19.9%	19.9%	19.9%	19.9%	19.9%
Dollar Investor Harm Predicted	55.7%	57.1%	55.7%	55.6%	54.6%
Investor Harm Predicted at IH-Score cutoff =1					
True Positives					
# Investor Harm Events	2,656	2,656	2,656	2,656	2,656
Investor Harm Events Predicted	1,889	1,966	1,889	1,882	1,883
True Positive (%)	71.12%	74.02%	71.12%	70.86%	70.90%
False Positives					
# No Investor Harm	1,012,217	1,012,217	1,012,217	1,012,217	1,012,217
No Investor Harm False Positives	334,126	323,049	334,393	331,758	332,190
False Positive (%)	33.0%	31.9%	33.0%	32.8%	32.8%
Proportion of $ Harm Predicted	73.5%	77.2%	73.5%	72.3%	73.4%

Table 6: Sensitivity to Including Additional Sets of Information to BrokerCheck

#	Model Specification	Investor Harm Predictions at IH-Score cutoff =1	
		Proportion of Harm Events predicted	Proportion of Dollar Harm predicted
	Baseline	71.1%	73.5%
	Additional Information Added to Baseline:		
1	Undisclosed Financial Events	71.1%	73.5%
2	Undisclosed Disciplinary Events	70.9%	72.3%
3	Exam Performance	70.9%	73.4%
4	Undisclosed Financial Events + Exam Performance	70.8%	73.4%
5	Undisclosed Disciplinary Events + Exam Performance	71.0%	73.4%
6	Undisclosed Financial Events + Undisclosed Disciplinary Events	70.9%	72.3%
7	Undisclosed Financial Events + Undisclosed Disciplinary Events + Exam Performance	70.9%	73.4%
8	HAC	74.0%	77.2%
9	HAC + Undisclosed Financial Events	73.9%	77.1%
10	HAC + Undisclosed Disciplinary Events	73.9%	77.0%
11	HAC + Exam Performance	74.4%	77.1%
12	HAC + Undisclosed Financial Events + Exam Performance	74.4%	77.1%
13	HAC + Undisclosed Disciplinary Events + Exam Performance	74.3%	77.1%
14	HAC + Undisclosed Financial Events + Undisclosed Disciplinary Events	73.9%	77.0%
15	HAC + Undisclosed Financial Events + Undisclosed Disciplinary Events + Exam Performance	74.3%	76.9%

Table 7: Impact of Replacing Financial and Disciplinary Disclosures on BrokerCheck by HAC

	Investor Harm Predictions at IH-Score cutoff =1	
	Proportion of Harm Events predicted	Proportion of Dollar Harm predicted
Baseline	71.1%	73.5%
Baseline - Disclosed Financial Events + HAC	73.9%	77.2%
Baseline - Disclosed Criminal Events + HAC	73.8%	77.3%
Baseline - Disclosed Disciplinary Events + HAC	73.6%	76.9%
Baseline - Association with Expelled Firm + HAC	73.7%	76.9%

Appendix A: Description of Broker Characteristics

#	Characteristic	Description of Characteristic
1	Complaint Settlements and Awards	Total number of Customer Complaints that led to settlement (above relevant threshold) or awards against the broker (since first registration) until the year under consideration
2	Judgments and Liens (Unsatisfied)	Total number of judgments and liens that are not satisfied through the year under consideration
3	Judgments and Liens (Satisfied)	Total number of judgments and liens that are satisfied through the year under consideration
4	Bankruptcy Disclosures (< 10 years)	Number of Bankruptcy disclosures that occurred within the last 10 years of the year under consideration
5	Bankruptcy Disclosures (> 10 years)	Number of Bankruptcy disclosures that occurred more than 10 years prior to the year under consideration.
6	Disclosed Disciplinary Events	Total number of regulatory actions, investigations, civil judicial actions, and terminations that are included on BrokerCheck through the year under consideration
7	Undisclosed Disciplinary Events	Total number of closed or dismissed regulatory actions, investigations, and civil judicial actions, and internal review that are not included on BrokerCheck through the year under consideration
8	Criminal Events	Total number of criminal disclosures through the year under consideration
9	Exams Passed	Exams (S6, S7, S63, S66) passed through the year under consideration
10	Exams Failed	Number of times exams (S6, S7, S63, S66) failed through the year under consideration
11	Average Exam Scores	Cumulative average of Series 6, 7, 63 and 66 exam scores through the year under consideration
12	HAC	Average number of investor harm events per registered rep (RR) for all other RRs at the same firm, averaged over the last five years across all firms the broker is employed by in the year under consideration
13	Association with Expelled Firm	An indicator that equals 1 if the broker was registered with any firm that was previously expelled
14	Number of Prior Employers	Number of prior employers through year under consideration
15	Employment Years	Number of years registered with FINRA through year under consideration
16	Dual Registration	An indicator that equals 1 for brokers who were registered with the SEC as investor advisors
17	Gender (Male)	An indicator for broker gender. The indicator equals 1 for male brokers and 0 for female brokers
18	Market Return	Annual return on S&P500 Index