Tone at the Bottom: Employee Chatter as an Indicator of Misconduct Risk

Dennis Campbell
Harvard Business School
dcampbell@hbs.edu

Ruidi Shang
Tilburg University
r.shang@uvt.nl

29 Nov 2019

Abstract: This paper examines whether "employee chatter", extracted via text-based statistical methods applied to comments left on the company review site Glassdoor, can be used as an indicator of misconduct risk. We argue that inside information on the incidence of specific acts of misconduct, as well as the control environments and broader organizational cultures that contribute to its occurrence, are likely to be widespread among employees. Our results show that this information can be used to develop measures that have potentially useful properties for measuring misconduct risk: namely that they increase prior to realizations of misconduct risk, decrease during periods of enforcement and likely remediation, and are not simply proxies for more readily available aggregate reviews. Importantly, in out of sample tests, they are also useful in predicting realized misconduct risk above and beyond other readily observable characteristics such as firm size, performance, industry risk, prior violation history, and overall employee ratings.

Tone at the Bottom: Employee Chatter as an Indicator of Misconduct Risk

I. Introduction

From 2008 to 2017, there were more than 26,000 violations of laws or regulations committed by public firms in the U.S., which incurred over 300 billion dollars of penalties (Good Jobs First 2019). The severity of corporate misconduct, and the variety of organizational factors that can give rise to it, raises important questions about how to measure and assess these risks. Recent work suggests that compared to external investigators such as investors and regulators, employees have information advantages about corporate misconduct given that they can observe or even experience it directly as a by-product of their daily work (Dyck, Morse, and Zingales 2010). Despite these observations, little is known about the type of the information that lower-level employees have about control risks in their organizations or how such knowledge can be surfaced. This paper examines whether ‘employee chatter’, extracted via text-based statistical methods applied to comments left on the company review site Glassdoor, can capture such information and be used as an indicator of misconduct risk.

We argue that inside information on the incidence of specific acts of misconduct, as well as the control environments and broader organizational cultures that contribute to its occurrence, are likely to be widespread among employees. Unlike financial disclosure or conference calls, it is difficult for external stakeholders (e.g., investors and regulators) to obtain inside information about firms’ operations, culture, and internal control environments. The emergence of company review websites provide external stakeholders with an information source to get access to employees’ views about their firms. In this study, we rely on data from Glassdoor.com to develop our text-based measures of employee chatter. Glassdoor is one of the most well-known company review websites. It has a rich coverage of approximately two million reviews on more than 5,000 U.S. listed firms from 2008 to 2017. To measure corporate misconduct, we draw on corporate violations from Violation Tracker, a search engine that covers corporate violations in the U.S. since 2000.

These data sources allow us to develop our analysis in two stages. First, we follow textual analysis approaches like those used in Taddy (2013), and Gentzkow Kelly, and Taddy (2017) which allow us to develop measures of employee "chatter" related to misconduct risk by first learning which words tend to occur in reviews differentially for firm-years in which violations occurred versus otherwise similar firm-years with no violations. Next, we use these measures, along with machine learning techniques to predict future violations and violation related outcomes like assessed penalties. We address whether such measures of employee chatter are useful indicators of misconduct risk by gauging the extent to which they are ultimately useful in predicting the actual realizations of such risk, namely, corporate violations.

Our results suggest that the measures of employee chatter we develop have potentially useful properties for measuring misconduct risk: namely that they increase prior to realizations of misconduct risk, decrease during periods of enforcement and likely remediation, and are not simply proxies for more readily available aggregate reviews. Perhaps most importantly, they are useful in predicting realized misconduct risk above and beyond other readily observable characteristics such as firm size, performance, industry risk, prior violation history, and overall employee ratings. Our findings collectively also point to evidence that our measures of employee chatter may be most useful when predicting longer term misconduct risk, particularly in samples of "good" firms with little prior misconduct history.

Our findings contribute to the broader literature on corporate misconduct and management control. While most prior studies focus on how to detect or control accounting fraud and misreporting (Brazel et al. 2009; Dechow et al. 2011; Purda and Skillicorn 2015), there is limited evidence about how to assess and control the risks of other forms of misconduct in firms. Early research documents the association between corporate misconduct and various firm features, including firm size, industry, and history of prior misconduct (Baucus and Near 1991). Recent evidence also documents the link between workplace safety and managers' needs to meet analysts' forecasts (Caskey and Ozel 2017). Further, Heese and Perez-Cavazos (2019) document the effects of headquarters' visits on facility-level misconduct. Whereas these studies examine market pressure and firm characteristics as determinants of misconduct and managerial monitoring of employees as a way to mitigate it, we focus instead on understanding the information that employees themselves have about broader misconduct risk in their firms and how this information can be captured to develop better measures of misconduct risk.

Our study also complements the literature on whistle-blowers. Dyck, Morse, and Zingales (2010) suggests that compared to external actors, employees have better access to information on corporate misconduct. Moreover, Whistleblower complaints are typically backward looking, happening when specific acts of misconduct have already occurred, and whistleblowers often bear considerable personal costs. Thus, the number of employees willing to produce such information is likely to be small relative to the total number of employees who have information relevant for assessing misconduct risk. Our study highlights the possibility of obtaining broader employee information from external reviews and documents that useful measures of "employee chatter" can be extracted and used to assess the risk of corporate misconduct.

We believe that our findings also have practical implications in light of the considerable financial, legal, and reputational risks that can arise due to corporate misconduct. Our results suggest that approaches like the one we take in this paper for measuring employees' inside information may be useful in developing leading indicators of the quality and state of an organizations internal control environment or even broader organizational culture which would otherwise be difficult for interested outsiders (e.g. investors, regulators, law enforcement, or even boards of directors) to observe.

The reminder of this paper is structured as follows. Section II reviews and discusses the related literature. Section III describes our data and primary variables. Section IV presents our methodology and results. Section V concludes the paper.

II. Literature and Theory

Corporate Misconduct

"Corporate misconduct" refers to the acts of employees or managers that violate rules, regulations, or laws (Vaughan 1999). Numerous cases of corporate misconduct have emerged globally in recent years and caused large financial and non-financial losses for organizations, their stakeholders, and even broader society (Association of Certified Fraud Examiners 2016, 2018). For example, Wells Fargo was caught in 2016 for opening two million fake accounts and selling products and services to customers under false pretenses to increase sales figures. More than five thousand employees were dismissed, and Wells Fargo paid an extremely high price, including billions of dollars spent on fines, settlements, and other costs, a depressed stock price, and long-term damage to its reputation (Srinivasan et al. 2017). More broadly, from 2008 to the end of 2017, there were more than 26,000 violations of laws or regulations committed by public firms in the U.S., which incurred over 300 billion dollars of penalties (Good Jobs First 2019). The severity of corporate misconduct brings an important question to investors, regulators, management, and boards of directors: how can these risks be detected, and mitigated, before they materialize?

Corporate misconduct may occur at both the operational (e.g., violating regulations about customer protection or employment) and management levels (e.g., violating accounting standards or tax regulations). Most prior studies investigating corporate misconduct focus on misconduct that occurs at the management level. Such misconduct tends to be concentrated by involving one or a few higher-level managers and their direct subordinates. A typical example of management-level misconduct is accounting fraud and misstatement. Purda and Skillicorn (2015), for example, show that the language used in the management discussion and analysis (MD&A) section of financial reports can be used to detect accounting fraud. Evidence also indicates that managers' voice in conference calls as well as the characteristics of senior managers and the board of directors is associated with firms' risk of accounting fraud (Hobson et al. 2012; Uzun, Szewczyk, and Varma 2004; Jia, Ven Lent, Zeng 2014; Liu 2016).

While most prior studies focus on management-level misconduct, and especially accounting fraud and misstatements, there is less evidence about how to assess the risks of operational-level misconduct. Such misconduct is typically committed by lower-level managers and employees, and can be more widespread than managerial misconduct. For example, the sales misconduct in Wells Fargo (i.e., creating unauthorized accounts) was largely committed by branch workers and managers in their daily practices, and these acts were widespread across branches nationwide. Information in financial reports and conference calls is less likely to be useful in assessing the risks of this type of misconduct, given these information sources usually do not contain details of firms' operational and managerial practices.

Depending on who is held accountable, operational-level misconduct can be studied in at least two ways. The first approach takes the view of management control and examines how managers can hold employees accountable for misconduct through their control decisions and practices. For example, Heese and Perez-Cavazos (2019) document that headquarter managers' visits to lower-level facilities can reduce the frequency of facility-level misconduct as well as the associated penalties. Other studies also document that managers can control employee misconduct through their design of control mechanisms, such as compensation and performance transparency (Chen and Sandino 2012; Maas and Rinsum 2013).

In contrast, the second approach examines how other stakeholders (e.g., employees, investors, regulators) can hold managers accountable for misconduct occurring in their organization or division. Managers' and employees' inherent tendency to commit misconduct is deeply rooted in the control environment of their organization (Wilks and Zimbelman 2011; Liu 2016). Managers are responsible for shaping the control and operating environment in their organization/division through their control decisions and practices. For example, even though the misconduct in Wells Fargo was committed by lower-level employees, their motive to engage in such misconduct was derived from the aggressive sales targets and cross-selling strategies set by middle- and upper-level managers. Further, the weaknesses in Wells Fargo's internal control systems also provided employees with opportunities to engage in misconduct.

Following this approach, information about the control and operating environment in organizations may be useful in assessing the risk of corporate misconduct. However, unlike financial reports and conference calls, the inside information about an organization's control and operating environment is less accessible to outsiders such as investors, regulators, and auditors. Employees, on the other hand, have the best access to such information as a by-product of their normal work (Dyck, Morse, and Zingales 2010). In their daily practices, employees can observe or even be involved in problematic practices which would eventually lead to misconduct. Therefore, the information provided by employees about the control and operating environment of their organizations may be useful in assessing the risks of corporate misconduct, especially those occurring at the operation level.

Possessing such information, however, does not necessarily lead to employees sharing their inside information about actual or potential corporate misconduct. Whistle-blowing, for example, is an important channel to obtain inside information about corporate misconduct risk from employees. Evidence indicates that employee whistleblowing allegations can deter corporate misconduct due to the threat of increased monitoring (Wilde 2017; Wiedman and Zhu 2018). However, a key limitation of using employees' whistle-blowing as an information source is the high social and economic costs faced by whistleblowers. Many employees choose to stay silent to avoid these costs (Dyck, Morse, and Zingales 2010; Rapp 2010, 2012). Another limitation of whistle-blowing is that the information revealed by employees is backward-looking. That is, when a misconduct is revealed by a whistle-blower it has usually progressed to a severe stage. Given these limitations, it is important to find out whether there are other channels for employees to share their information about the control and operating environments of their organizations.

Employee Chatter

We address this question by examining whether employee chatter extracted from their reviews about their firms can be used to assess the risk of corporate misconduct. The emergence of social websites provides employees with platforms to talk about their firms. There are several websites (e.g., Glassdoor, Great Place To Work, Indeed) which provide employees with opportunities to review their former or current employers. Compared with other information channels such as survey and whistle-blowing programs, these websites have several advantages. First, the anonymity of these websites significantly reduces the social and economic costs faced by employees who share negative information about their firms. Second, the information shared by employees on these websites is not limited to corporate misconduct, but rather encompasses broadly the general observations and experiences of employees in their firms. In other words, this information may not only be used to detect any ongoing misconduct, but can also reveal any problematic practices that could lead to future misconduct. Further, due to the popularity and easy access of these review websites, they attract a large number of users and collect enormous numbers of employee reviews. In other words, these websites are likely to contain rich information about firms' internal control and operating environments, which is otherwise difficult for outsiders to obtain through other channels.

The Glassdoor website (hereafter, Glassdoor) is the leader in this firm review segment. In Glassdoor, employees can rate their firm in several perspectives and write down their opinions about the pros and cons of the firms. The comprehensive coverage, rich amount of reviews, and well-structured format of Glassdoor make it a useful source for examining employee comments. In particular, Glassdoor launched its firm review feature relatively early (i.e., since 2008) and collected over 1.7 million reviews of more than 5,000 public firms in the U.S. until 2017. The large amount of employee reviews on Glassdoor makes it a potential useful source to obtain inside information about firms. Second, employees have incentives to share information on Glassdoor, as they need to provide at least one firm review to get access to the information posted on Glassdoor, including company reviews as well as information about new positions and interviews. Third, Glassdoor provide a clear and specific structure for employees to share their reviews. When sharing their reviews, employees are asked to rate their firm in several different dimensions and write down the pros and cons of their firm as well as their advice to the management team. This structure helps control the quality of employee reviews, making each review more informative and understandable to others.

Several studies have investigated the information on Glassdoor. For example, Hales et al. (2018) suggests that the outlook ratings that employees gave their firms on Glassdoor predicts financial statement line items (including sales, gross margin, earnings, restructuring expenses, impairments, and write-downs) earnings surprises, as well as management guidance. Evidence indicates that the outlook ratings can also predict stock returns (Green et al., 2017) and analyst output (Huang et al., 2017). While the findings of these studies support that employee reviews on Glassdoor are likely to contain important information about firms, focusing on the ratings is subject to several limitations. First, the categories of those ratings are homogenous for all the firms on Glassdoor. In other words, the ratings do not capture firm-specific information that is not covered by these categories. Second, without specific guidelines or definitions, employees may have different interpretations about what each category means and covers. That is, the information captured by the ratings is likely to be noisy. Finally and most importantly, the five-point scale of the ratings only captures employees' overall evaluation of the given perspectives, but does not contain specific information about why employees chose to give a certain point. Different employees may experience a firm very differently, and they may also weigh the positive and negative aspects they experienced in different ways. However, the ratings are overly aggregated and not informative about any of these details.

In comparison, the written comments about the pros, cons, and advice to management are likely to contain information that is detailed and specific to each firm. Another advantage of the written comments is that Glassdoor separates the pros, cons, and advice sections. A word used to describe the pros of a firm could mean something very different when it is used to describe the cons. For example, when the word "ethical" is mentioned in the pros (cons) of a firm, it is likely to mean the firm has ethical (unethical) practices. Using the words extracted from the written comments on Glassdoor, we investigate whether employee chatter can be used to assess the likelihood of corporate misconduct.

III. Data, Sample, and Variables

Data and Sample

We use data from three sources for this study. First, we collect data on employee comments from the Glassdoor website. The Glassdoor website collects people' reviews about their current or former employers since 2008, and displays the reviews anonymously to its users. Registered users can write comments about the pros and cons of their firm, give advice to the management, and rate the firm in several dimensions. We extract over 4,000,000 reviews of more than 50,000 U.S. organizations from June 2008 (when Glassdoor first launched the review section) to March 2017. We focus on public firms only and leave out firms that received less than ten reviews during this period (i.e., on average less than one review per year). Second, we obtain firm data (e.g., size, capital structure, profitability) from Compustat. Our final sample for which we have data on firm characteristics, Glassdoor ratings, and sufficient Glassdoor review comments consists of 10,156 firm years representing 1,388 unique firms during the period 2008-2017.

Third, we obtain the data on corporate misconduct from Violation Tracker, a search engine that covers civil and criminal cases brought against firms. The data is produced by the Corporate Research Project of Good Jobs First. Violation Tracker covers "banking, consumer protection, false claims, environmental, wage and hour, unfair labor practice, health, safety, employment discrimination, price-fixing, bribery and other cases initiated by 43 federal regulatory agencies and the Justice Department since 2000" (Good Jobs First, 2019). For completeness, Violation Tracker complements agency enforcement records with information collected on settlements announced in press releases. Joint ventures in which a parent company owns more than 50% are treated as owned facilities; otherwise they are treated as independent companies. Considering the importance and influence of the violations, Violation Tracker removes violations where the penalty or settlement is lower than $5,000. We extract all 26,934 violation cases committed by public U.S. firms over the period of 2008 and 2017 on Violation Tracker. Most of these cases were committed by "repeat" violators, that is, firms committing multiple violations within a year or/and repeatedly committing violations throughout multiple years.

Variables

Corporate misconduct

We measure corporate misconduct based on firms' violations of rules and regulations issued by government institutions. Focusing on violations as our measure of misconduct has several advantages. First, violations are an observable measure of the occurrence of misconduct, and the underlying undesired behaviors that cause them are likely to be significant for both organizations and their stakeholders. Second, violations encompass many forms of misconduct, allowing us to assess misconduct risk across a variety of firms and industries. Further, violations usually have significant consequences for firms in terms of fines, regulatory scrutiny, damages of reputation, and legal risk. These consequences make violations and their causes an important focus of managers, investors, and regulators.

We first classify violations into two types: those that are directly related to workers or consumers and thus more visible to employees,[1] and those that are more likely to occur at the management level and thus less visible to employees.[2] Our intuition is that employees are likely to have more information about misconduct that occurs in their daily operations and practices than those occurring at the management level of their organization. For each violation case,

[1] For example, the violation of customer protection, safety, health or employment regulations.
[2] For example, accounting deficiencies and fraud, bribery, kickbacks, money laundering, and tax fraud.

Violation Tracker provides information about the reason for offense and identifies the agency that is responsible for any related enforcement actions. We consider a violation to have high visibility for employees if its stated reason for offense includes one or more of the following words: safety, labor, wage, employ/employment, workplace, and consumer/customer; otherwise, we classify it as low-visibility. The Appendix lists all the 67 different types of violations appearing in our sample, and the enforcement agencies that are involved as well as our classification on the dimension of high and low visibility.

Most of our tests use aggregate measures of violations including the occurrence of a violation, the number of violations, and assessed penalties. For each firm-year observation, we construct (1) two dummy variables indicating the presence of high-visibility and low-visibility violations (*ViolationHV* and *ViolationLV*), respectively; (2) two discrete variables capturing the number of each type of violations (#*ViolationHV* and #*ViolationLV*); and (3) the penalties imposed by the relevant regulatory or legal authorities on the perpetrating firm (*Penalty*). We take the view that factors such as high pressure incentive systems, failures in internal controls, and broader corporate cultures that encourage individuals to take actions contrary to stated policies and values are the ultimate sources of misconduct risk in organizations. Moreover, while the primary factors that drive it remain similar across organizations, misconduct risk can manifest in different ways depending on industry and firm characteristics. We take these aggregate measures as overall proxies for the realization of such risk, and we control for underlying heterogeneity in types of violations and opportunities to commit them, through the use of industry fixed effects and firm characteristics such as firm size, profitability, and lagged violation.

Of course, the use of violations as our proxy for misconduct has several limitations that are worth considering. First, violations are a downstream measure of both potential and realized misconduct risk. Violations and related penalties are recorded in the data at the time of enforcement action, not necessarily when the actual underlying misconduct takes place. This makes it difficult in practice to approximate the appropriate lead-lag structure in our empirical analyses. We attempt to account for this in our analyses by using both short-lag and longer-window estimation approaches. We also note that violations relating to workplace health and safety, wages and hours, and labor relations are among the most frequent types of violations in our data and would likely involve a relatively short lag between realized misconduct risk and enforcement actions (Heese and Perez-Cavazos 2019).

Second, we only observe a measure of misconduct for those firms that are subject to an enforcement action. We do not observe risk realizations for firms that commit acts of misconduct that are not detected, or are detected and ultimately not enforced by the relevant legal or regulatory institutions. In practice, this means that we may fail to detect many cases of underlying misconduct. While we note this as a general limitation, this is a well understood problem in studies of corporate fraud and compliance risk and is not unique to this paper (Soltes 2019). As we discuss in the following sections, we believe our approach in this paper has promise for measuring underlying potential risk of misconduct based on extracting useful signals from ongoing employee reviews on popular platforms such as Glassdoor.

Employee chatter

We measure employee chatter using employees' comments about their current (or former) firms, including the words employees used to describe the pros and cons of their firm as well as the advice they gave. We take several steps to clean and organize the words used in those comments. First, we eliminate any non-English words, numbers, punctuations, stop words (e.g., “the”, “is”, “at”, “that”), and spelling mistakes that cannot be fixed.[3] Second, we replace each word with its root using the Porter stemming algorithm (Porter 1980) and the Natural Language Toolkit (Loper and Bird 2002) to reduce redundancies. Third, we drop words that are used in more than 50% reviews or less than five reviews. Our rationale for taking this step is that words that are too common (i.e., used in more than half of all the reviews) or too rare (i.e., used in less than five reviews) are less useful in assessing firms’ misconduct risks. We end up with a vocabulary of 11,233 unique words for our sample firms.

Our primary goal in using this vocabulary is to see if we can extract useful signals of broader corporate misconduct risk from employee reviews of firms. The basic idea as outlined earlier is that employees may have inside information on both the incidence of specific acts of misconduct as well as the control environment and broader organizational culture that contribute to its occurrence. Moreover, it may be possible to extract some of this information from the types of words that employees use to describe their firms.

To this end, we follow textual analysis approaches like those used in Taddy (2013), and Gentzkow Kelly, and Taddy (2017) which use inverse regression to project text onto the outcomes of interests. These techniques allow us to obtain low dimensional document scores while still preserving information contained in the reviews.[4] Using this approach, we develop measures of employee “chatter” related to misconduct risk by first learning which words tend to occur in reviews differentially for firm-years in which violations occurred versus otherwise similar firm-years with no violations. We then assign weights to each word in our vocabulary

[3] Misspelling and mistakes are detected and fixed through three steps: (1) check if the word exists in the English dictionary, if yes, leave it as its current form; (2) if the word doesn't exist, account for corrections (i.e., any alphabet replaced, added or omitted). If corrected word exists in dictionary, leave it as the corrected form; and (3) if the corrected word still does not exist in the dictionary, it is likely to be a non-English word and therefore removed from our analysis.

[4] In our paper, a “document” refers to all Glassdoor reviews for a given firm-year. That is, we score each firm year based on all the words in our vocabulary of 11,233 words that appear across Glassdoor reviews for that firm-year.

based on how strongly they differentiate the violation from non-violation firm-years. Finally, we apply these weights to create a weighted sum of firm-year word-count relative frequencies based on all Glassdoor reviews for a given firm in a given year.

The resulting weighted sum(s) constitute our text-based measure(s) of employee chatter related to potential corporate misconduct. We construct such measures separately for the pros, cons, and advice sections of Glassdoor reviews to account for the fact that similar words may have different meanings in the context of the review section in which they appear. For example, the word "ethics" may have a different meaning when appearing as a con rather than a pro. Our methodology for constructing these measures is described in more detail in section IV below.

Firm Ratings

Besides writing comments, users of Glassdoor can give firms an overall rating and five specific ratings on firms' work-life balance, culture and values, career opportunities, company benefits and senior management. Both the overall and specific ratings are scales ranged from one star (i.e., the worst) to five star (i.e., the best). Recent work shows these ratings are associated with stock returns, earnings surprises, financial statement line items, and analyst outputs (e.g., Green et al., 2017; Hales, Moon, and Swenson, in press; Huang, Li, and Markov, 2017). In this study, we examine whether the words used in employee comments have incremental predictive power after controlling for these ratings; and if so, whether the words have higher, or lower, predictive power than the ratings.

Firm Characteristics

We control for firm size, leverage, profitability, lagged violations, and industry fixed effects in our analyses, given that prior research suggests these characteristics may be associated with corporate misconduct (Baucus and Near 1991). First, larger firms usually have more sub-units and employees, higher degrees of decentralization, and higher levels of information asymmetry. The effectiveness of monitoring systems also tend to decrease as a firm grows. These features increase the opportunities for misconduct to occur in large firms (Finney and Lesieur, 1982; Vaughan, 1982). We control firm size (*Size*) using the natural logarithm of firms' total assets. Second, we control firm leverage (*Leverage*) using the debt-to-equity ratio, as a firms' capital structure may pressure them to enhance their performance in ways that increase the risk of misconduct. Additionally, low profitability may also pressure firms to find alternative sources of resources or improve efficiency in unethical and/or illegal ways (Baucus and Near 1991). We measure and control firms' profitability based on their return on assets (*ROA*). Further, a history of prior misconduct indicates that a firm may be subject to an unethical culture which could lead to repeated misconduct in the future (Baucus and Near 1991). Therefore, we control firms' lagged violations of rules and regulations (*Lagged_outcome*) in our analyses. Moreover, misconduct is more frequent in certain industries than in others; surveillance and legal enforcement are also stricter in some industries than in others (Baucus and Near 1991). In order to address the underlying differences in industry risk and enforcement, we include industry fixed effects in our analysis.

Descriptive Statistics

Summary statistics for our sample are provided in Table 1. The first observation worth noting is that violations are relatively infrequent. Those we classify as highly visible occur in 17.8% of firm-years on average while low visibility violations are about half as frequent at only 9.3%. Despite the majority of firms having no violations in a given year, the number of high visibility violations averages 45.3 with a relatively large standard deviation. This is due to a relatively small number of firms which are penalized for many individual violations in a given year. These firms tend to be concentrated in industries such as railroad transportation, energy and mining, and air transportation. We retain these firms in our data since most of our analyses effectively collapse these down into aggregate measures such as indicators for the occurrence of a violation or aggregate penalties. However, we note that the thrust of our results remains substantively similar when the small number of firms with such outlier observations are excluded from our analyses.

Aggregate employee ratings on Glassdoor appear to have a relatively narrow distribution around a mean of 3. However, the underlying number of reviews and overall number of words in employee comments varies widely across firm-years in our sample. For this reason, it is important in most of our analyses that we normalize counts of specific words in our vocabulary by the total number of words contained in employee comments for a given firm-year. That is, for most of our primary analyses, we examine relative word frequencies rather than raw word counts.

Finally, firms in our sample, perhaps not surprisingly, demonstrate significant heterogeneity in characteristics like size, leverage, and return on assets, all of which for reasons noted above may be correlated with misconduct risk. To assess whether our text-based measures of employee chatter are useful in predicting misconduct, it is important not only that they correlate with risk but that they do so incrementally relative to other potential indicators of misconduct risk. This is especially true of indicators like firm characteristics and past violation history which are readily observable to external and internal parties that may be interested in assessing such risk including investors, regulators, and boards of directors.

[Insert Table 1 Here]

IV. Empirical Methods and Results

Measuring Employee Chatter Related to Corporate Misconduct

We develop our text-based measures of misconduct risk using the following approach. For each word in our vocabulary, we estimate the following poisson regressions during the period 2008-2011 (i.e., the first three years of our sample period) in order to learn what words tend to occur differentially for firm-years with and without violations:

$$E(W_{jit}|x_{it}, v_{it}) = e^{\alpha_j + \beta_j x_{it} + \sum_{k=1}^{k=K} \phi_{kj} v_{kit} + \varepsilon_{it}} \tag{1}$$

Where W_{jit} denotes the count of word *j* across reviews for firm *i* in year *t*; x_{it} is a vector of controls for firm *i* in year *t* (Glassdoor ratings, size leverage, ROA, and industry and year indicators); and v_{kit} is a dummy for the presence/absence of violation type *k* for firm *i* during year *t*. We split violations into those with high and low visibility. Thus, K=2, and we include indicators for both types of violations in our specification. We estimate these regressions separately for counts of words from each of the three primary areas of Glassdoor reviews: pros, cons, and advice. This yields 33,699 separate regressions (3 comment categories x 11,233 words). We use data from the period 2008-2011 as a sort of "training set" since we will ultimately use information from this step along with forward regression to predict violations during a future period, namely 2012-2017.

We next use the estimates from (1) to calculate employee chatter indices of the form $\phi_{k1}\frac{W_{1it}}{\sum W_{it}} + \phi_{k2}\frac{W_{2it}}{\sum W_{it}} + \cdots + \phi_{k,11,233}\frac{W_{11,233it}}{\sum W_{it}}$ for firm *i* in year *t*, where $\sum W_{it} = \sum_{j=1}^{11,233} W_{jit}$ is the total count of all words appearing in reviews for firm *i* in year *t*. We conduct this transformation, and measure relative word frequencies, separately for words appearing in each of the pros, cons, and advice sections of the reviews. These transformations, thus, create six different weighted sums of relative word frequencies, one each for low and high visibility violation types within each of the three comment categories of pros, cons, and advice. These weighted sums provide a reduced form mapping of text onto violations and can be applied to any firm in any year where we observe review text, independent of whether or not a violation has yet occurred.

Note that rather than doing two regressions (one for each violation type) with 33,699 independent regressors (11,233 relative word frequencies x 3 review categories) which would be inefficient and computationally expensive, this approach allows us to reduce the dimensionality of the problem down to three weighted indices for each of the two primary violation types. Taddy (2013) and Gentzkow et al. (2017) show that under the (admittedly strong) assumption of independence of word frequencies within a document, the approach above yields indices which are a sufficient reduction (SR) of the text in the sense that once we have the SRs, we can ignore text for the purpose of predicting future violations.

The weighted indices described above capture the extent to which reviews for a firm in a given year reflect what we refer to as "misconduct words". Higher values of a given index suggest current or former employees are using words associated with violations more frequently in their reviews of a firm. We refer to these indices throughout the remainder of the paper as our "misconduct word" indices, or *MW_IndexHV and MW_IndexLV* for high- and low-visibility violations, respectively. The resulting six indices serve as our primary measures of employee chatter as it pertains to misconduct risk.

Table 2 shows the top 25 words in the pros, cons, and advice comment sections of Glassdoor reviews which receive the highest weights in calculating our misconduct word indices. Columns 1-3 (4-6) are the words that load most highly for the presence of any high (low) visibility violation in a given firm-year. Columns 7-9 are the words that load most highly for serial violating firms, defined as firms that had violations in at least 2 years during the period 2008-2011, which is a subsample we focus on in some of our later analyses.[5]

For each of these outcomes, our methodology appears to select intuitive words that are plausibly related to misconduct or broader internal control risks such as “safety”, ‘quality”, “supervision”, “standard”, and “downside”. It also selects some less obvious words such as “able”, “help”, “ahead”, “promote”, and “advance”. These words may relate less directly to actual internal control breakdowns, and hence specific acts of misconduct, and more to broader organizational culture characteristics that might contribute to misconduct risk. Cultural characteristics of organizations with high misconduct risk might include lack of support from management (associated with words like “able” and “help”) or high degrees of internal competition for pay or career advancement (associated with words like “promote”, “advance”, and “compensation”).

It is important to note that many of the words that are highly weighted by our methodology need not be viewed as negative *per se*. We rely on the idea that certain words showing up relatively more frequently in reviews signal an increase in the underlying risk of misconduct, due either to employees witnessing others engaging in specific acts of misconduct that have yet to surface externally or reflecting characteristics of their organization’s culture that create a risk of future acts of misconduct. So, for example, increasing frequency of words like “leadership”, “fair”, “boss”, or “integrity” may reflect increasing concerns about these particular aspects of organizations that employees might otherwise view positively.

[5] In some our later analyses, we do longer window tests where we attempt to predict which firms that show a low propensity to violate in one time period go on to commit violations over multiple future time periods – e.g. which become “serial violators”. For these tests, we construct our misconduct word indices in one long period (2008-2011) and use those indices along with other variables to predict misconduct patterns in a future period (2012-2017).

Many of these features can be seen when looking at specific reviews. For example, consider the following comment taken from the "cons" section of an employee review of Wells Fargo in 2011 prior to widespread public knowledge of the underlying misconduct in its sales practices:

> *"Some **downsides** of **working** at Wells Fargo include their **sometimes** ridiculous expectations and high **sales targets**, as well as not valuing their **employees** as **people**. **Leadership** needs to realize that just because you focus on a **product** or service on one particular day of the week shouldn't mean if you sell that product the day before, it's useless."*

The bold words above appear in Table 2, that is, they rank as top 25 in our Misconduct Word Indices. This comment directly points out the underlying problem in Wells Fargo's practices, even though it was made two years before the LA Times expose and five years before the company was eventually fined by various regulators. Meanwhile, the "pros" section of the same comment mentioned several positive sides of Wells Fargo, including "benefit", "salary", and "opportunity and room for career advancement". Interestingly, the employee eventually gave Wells Fargo a rating of four (out of five). This is consistent with our argument that the ratings are overly aggregate and less informative than the written comments, as employees tend to weigh the positives and negatives of their firms based on their (unobservable) personal experiences and judgements.

Of course, not all of the top words in our index are obviously intuitive which may reflect the difficulty of predicting violations with underlying text in employee reviews. Whether or not our text-based measures of employee chatter are predictive of future misconduct is the central question in this paper.

[Insert Table 2 Here]

Before developing formal prediction models, it is useful to consider descriptive patterns in these measures for firms that are known to have engaged in misconduct over particular periods.

For illustrative purposes, Figure 1 shows an aggregate summary measure of our underlying misconduct word indices for Wells Fargo during 2008-2017, a period in which misconduct in its sales practices was reported to be widespread throughout the organization (Srinivasan et al. 2017). Two interesting patterns emerge. First, our summary index for Wells Fargo grows relatively steadily through 2013 to over 2.5 times its baseline value in 2008. In December 2013, the Los Angeles Times (LA Times) published what is reported to be the first major expose on potentially widespread fraudulent sales and account opening practices at Wells Fargo ("Wells Fargo's pressure-cooker sales culture comes at a cost, *Los Angeles Times*, December 21, 2013) . The LA Times article is broadly credited with bringing public awareness as well as legal, regulatory, and political scrutiny to the bank for its sales practices. The trend in our measure appears to track well with the period over which Wells Fargo was reported to be engaged in increasingly fraudulent sales practices, and it appears to lead external awareness by several years.

The second pattern of interest that emerges is that our measure then declines dramatically after 2013 and through 2017, a period in which Wells Fargo was reported to be engaged in several actions aimed at mitigating misconduct risk in its sales practices. These actions included, among others: reducing sales goals; conducting a detailed internal investigation led by an outside law firm; and firing employees, middle and senior management, and even the CEO (Srinivasan et al. 2017). Again, the trend in our measure appears to track well with the period in which Wells Fargo was reported to be engaged in broad organizational changes aimed at reducing the risk of misconduct in its sales practices.

[Insert Figure 1 Here]

While Wells Fargo serves as a single illustrative example, the same general pattern emerges when we consider a broader sample of firms that undergo a substantial change in realized misconduct risk. Figure 2a shows a summary misconduct word index with the same underlying calculation as in the Wells Fargo example above for two subsamples of firms, both of which had no violations over the period 2008-2011. “Non-violators” are those firms that also had no violations during 2012-2017 (N=654 firms). “Serial-violators” are those firms that went on to have violations in at least two of the years during the period 2012-2017 (N=126 firms). As in the Wells Fargo example in Figure 1, firms that transition to serial violator status show a relative increase in our index prior to 2012 with a decrease roughly back to baseline during the period in which enforcement actions eventually took place. While we have less direct information on this set of firms than on the widely publicized case of Wells Fargo, the pattern suggests that our measure is tracking well with periods where misconduct risk is high and decreasing during periods when such misconduct is penalized and firms are likely to be taking remedial actions.

[Insert Figure 2 Here]

Interestingly, Figure 1 and Figure 2a show that Glassdoor ratings remain relatively flat over the entire period both in the specific case of Wells Fargo and in the more general case of non-violator to serial-violator transition firms despite the significant changes in our misconduct word indices. Table 3 shows more formally that, while many pairwise correlations are negative and statistically significant, there are relatively low correlations between overall ratings and our various misconduct word indices. This is in part by design as we controlled for ratings in estimating the underlying word weightings for these indices in order to extract incremental information. More intuitively, high values of these indices and related increases in violation-associated words need not necessarily occur only when employees are dissatisfied with their firms. As discussed above, ratings aggregate many different dimensions of organizations and jobs that employees may value, and employees may even value features of their organizations that can also contribute to misconduct risk like ambitious goals, competitive work environments, strong pay for performance schemes, and freedom in decision-making.

[Insert Table 3 Here]

Figures 1 and 2 along with Table 3 provide descriptive evidence that our misconduct word indices have some potentially useful properties for measuring misconduct risk: namely that they increase prior to realizations of misconduct risk, decrease during periods of enforcement and likely remediation, and are not simply proxies for more readily available aggregate reviews. We turn our attention next to developing more formal models to predict misconduct outcomes.

Predicting Misconduct

Our approach to assessing the value of employee chatter as a measure of misconduct risk is to measure its influence in improving predictions of future violations. We predict violations using the misconduct word indices developed earlier in this section. For each of our major violation categories (high and low visibility), we estimate a gradient boosted tree, where the dependent variable is an indicator for the occurrence of a violation for a firm in a given year. In our initial analyses, we focus on one year ahead predictions of violations. In later analyses, we supplement these with longer window predictions of changes in violation status over multi-year periods. We randomly split firms in our sample into training and test sets, allocating 80% of firms (and their related firm-year observations) to the former and 20% of firms (and their related firm-year observations) to the latter. We estimate our prediction models on the training set and test the resulting predictions on the 20% of firms held out for the test set. We perform this predictive analysis only for the 2012-2017 period with the idea that we learned the appropriate weights for our misconduct word indices using information from 2008-2011 and are now using these weights to project text onto a measure of future misconduct risk (*MW_IndexHV and MW_IndexLV).*

The gradient boosting algorithm we use for prediction hews closely to the multiple additive regression tree (MART) algorithms described by Hastie et al. (2017) and Schonlau (2005) and proceeds roughly as follows:

1. Set initial predictions to the constant only model that minimizes a suitable loss function (e.g deviance in the case of logistic regression for violations)
2. For all regression trees m =1 to M
    a. Compute residuals based on the current model (or pseudo-residuals in the case of a binary classifier like *ViolationHV* or *ViolationLV*)
    b. Fit a regression tree with a fixed number of nodes to the residuals
    c. Compute the average residual for each terminal node of the tree
    d. Update by adding the regression tree of residuals to the current best fit

In effect, this iterative procedure estimates multiple regression trees, each of which fits the residuals of all previous trees combined (Schonlau 2005). In this way, the model learns and corrects for over or under-adjustments in predicted values at each iteration. We follow advice in Hastie et al. (2017) and Schonlau (2005) in two important ways when implementing our boosting algorithm. First, we control the size of the regression trees (number of nodes) by allowing up to 5-way interactions among the predictors in our model, with each interaction corresponding to a split in a regression tree. Second, we include a shrinkage parameter of 0.01 in step 2(d) and set M=10,000 iterations (e.g. 100/shrinkage parameter per Schonlau 2005). This is a form of regularization in which the shrinkage parameter controls the rate of learning in the model and reduces the impact of each successive regression tree on the updated estimates in step 2(d). The idea is that many small steps in the learning process (controlled via shrinkage and the number of iterations) can help avoid overfitting in the training data which might lead to poor predictions in the test sample.

Table 4 shows results from fitting a logistic model using this boosting algorithm, where the dependent variable is an indicator for the occurrence of a high or low visibility violation during year *t*. The predictors include firm characteristics (*Size, Leverage, ROA,* and an indicator for the occurrence of a violation in the prior year), indicators for industry and year, and data from Glassdoor (average overall rating and our misconduct word indices). Prediction performance is measured on the test sample of firms using the pseudo-R-squared, calculated as $R2 = (1\text{-}L1/L0)$ where $L1$ and $L0$ are the log likelihood of the full model and intercept-only models, respectively. The reported influence statistics measure the relative contribution of each variable to the improvement in the log likelihood across training iterations.

Table 4 shows that the psuedo-R-squareds in the test sample are similar for predicting both high and low visibility violations. More interestingly, while prior violations and firm size are the most influential individual predictors for both violation types, our misconduct word indices contribute both individually and collectively to prediction performance. Individually, they account for anywhere from 3.5% to 5.6% of the improvement in prediction performance depending on the section of review (pros, cons, or advice) and the type of violation (high or low visibility). Collectively they account for approximately 14% and 14.4% in the case of high and low visibility violations respectively. In all cases, whether viewed individually or collectively, our misconduct word indices seem to be more influential in improving prediction performance than aggregate employee ratings for which measured influence is in the neighborhood of 1%. Similar patterns emerge for workplace health and safety violations which are the most frequent violation type in our sample.

Overall, the evidence in Table 4 suggests that our measures of employee chatter are useful indicators of misconduct risk. They are both influential in improving predictions of future violations and they are incrementally so relative to other readily observable characteristics such as firm size, performance, industry risk, prior violation history, and overall employee ratings.

However, there is no absolute benchmark on what constitutes a high pseudo-R-squared making it difficult to assess overall prediction quality. To this end, it is intuitive to consider the relatively high frequency of non-violation in our data. A naïve model that always predicts no violation would yield a correct prediction approximately 82.2% and 90.7% of the time for high and low visibility violations respectively in our sample. Our model performs considerably better than this benchmark in the cases of both high and low visibility violations with correct prediction rates of approximately 94% in both cases. For health and safety violations, the naïve model would yield correct predictions 88.1% of the time compared to our model at 93.2%.

[Insert Table 4 Here]

Table 5 shows largely similar patterns to those of Table 4 when broader violation related outcomes are considered including the total number of violations, whether a firm is subject to more than one violation, and total penalties.[6] For these analyses, we repeat the gradient boosting algorithm described above with appropriate loss functions for poisson regression, multinomial logit, and OLS for each of these outcomes respectively. For these analyses, we aggregate across all violation types and report combined influence statistics for *MW_IndexHV and MW_IndexLV* within each of the three review sections of pros, cons, and advice. So, for example, the combined influence of these indices is 4.2% (5.7%) for the cons (advice) sections of Glassdoor reviews when predicting total number of violations in a given firm-year. The influence statistics show that while lagged outcomes account for the lion's share of improvement in the underlying

[6] For penalties, we focus on predicting Ln(1+*Penalty)* given the highly skewed nature of this variable.

predictive models in all cases, the same pattern as in Table 4 remains, namely that our measures of employee chatter are individually and collectively influential in predicting these outcomes above and beyond other readily observable firm characteristics.

[Insert Table 5 Here]

A limitation of our analyses so far was hinted at earlier in the paper where we noted that actual acts of misconduct, breakdowns in internal controls, or deteriorations in organizational culture could all lead enforcement actions, and hence violations and penalties, by well more than a year in some cases. To see this possibility, one only needs to look at Figure 1 in the case of Wells Fargo and Figure 2 for the more general case of serial violator transition firms.

To address this issue, we collapse our sample down to two longer window time periods, 2008-2011 and 2012-2017, and focus on the set of firms which had either no violations or no more than 1 year with violations in the first period. We refer to these firms as non-serial violators and firms that experience 2 or more years of enforcement acts as serial violators. We reconstruct our misconduct word indices in the same manner described in the first part of this section of the paper, except in the case we regress word frequencies (over the whole first period) on serial violator status in the first period to recover the appropriate weights for each word in our vocabulary. That is, we discover which words most differentiate serial and non-serial violators in the first period. We then use those weights to construct our misconduct word indices for non-serial violators to see if they can be used to predict which of those firms transition to serial violator status in the second period.

We think this sample of initial non-serial violators is interesting to focus on for several reasons. First, transitions to serial violator status for any of these firms could plausibly be due to failures in organizational control systems or shifts in culture. Our motivation for employee chatter as an indicator of misconduct risk centers on the idea that employees have inside information not only on potential acts of misconduct but also on these broader features of their organizations. If this is the case, then deteriorations in any of these factors should be picked up by our text-based measures which should then predict subsequent related outcomes like transitioning to serial violator status. Second, and as noted extensively, this approach should overcome some of the limitations of our shorter-window one-year-ahead predictions. Finally, this is an interesting sample since there would presumably be less prior (or at least recent) violation history to draw on in making an inference about these firms' misconduct risk and employee chatter may be particularly informative.

We use an 80% (20%) train (test) sample split for these firms and use the same gradient boosting approach described above to predict transitions to serial violator status in the period 2012-2017. The predictors include firm characteristics (*Size*, *Leverage*, *ROA*, industry indicators, and an indicator for the occurrence of any violation in the period 2008-2011); and data from Glassdoor (average overall rating and our misconduct word indices). Both firm characteristics and Glassdoor ratings are taken in the model at their respective means over the 2008-2011 period.

The results of our prediction estimates are provided in Table 6. Two noteworthy results are apparent in this table. First, our model performs much better for predicting high versus low visibility violations. The test sample pseudo-R-squareds are 35.2% and 10.5% for high and low visibility violations respectively. A similar result emerges when considering the proportion of correct predictions in each case. The boosting algorithm yields similar correct classification rates of 90.1% and 91.1% respectively for low and high visibility violations. However, a naïve model that predicted no firms would transition to serial violator status over this period would yield corresponding correct classification rates of 93% for low visibility violations and 87.9% for high visibility violations. Thus, our boosting procedure only appears to improve on the naïve model for the latter suggesting that employees likely have considerably more information about the long term risk of these types of violations. The second noteworthy pattern is that our employee chatter measures are relatively highly influential in improving prediction performance in this sample, particularly for high visibility violations where they individually exceed even those of firm characteristics like size in some cases. Contrasting these results with those in Table 4 suggests that our measures of employee chatter may be more predictive of longer term misconduct risk, particularly in samples of "good" firms with little prior misconduct history.

[Insert Table 6 Here]

V. Conclusion

This paper examines whether "employee chatter", which is extracted from employee comments on the company review site Glassdoor, can be used as an indicator of misconduct risk. We argue that inside information on the incidence of specific acts of misconduct, as well as the control environments and broader organizational cultures that contribute to its occurrence, are likely to be widespread among employees. The results outlined in this paper show that it is possible to extract some of this information from the types of words that employees use to describe their firms on external review platforms like Glassdoor.

Moreover, our results show that this information can be used to develop measures that have potentially useful properties for measuring misconduct risk: namely that they increase prior to realizations of misconduct risk, decrease during periods of enforcement and likely remediation, and are not simply proxies for more readily available aggregate ratings. Importantly, they are also useful in predicting realized misconduct risk above and beyond other readily observable characteristics such as firm size, performance, industry risk, prior violation history, and overall employee ratings. Intriguingly, this may point to the utility of approaches like the one we take in this paper for measuring leading indicators of internal control breakdowns or degradation in organizational culture which would otherwise be difficult for interested outsiders (e.g. investors, regulators, law enforcement, or even boards of directors) to observe on a timely basis.

Our study contributes to the broader corporate misconduct and management control literatures. While other studies have examined the monitoring role of managers and boards in surfacing and mitigating misconduct, these roles rely on observing behavioral or other indicators of misconduct risk at successively lower levels in the organization. Less is known about the type of information lower level employees have about control risks in their organizations. Our measures of employee chatter and their link to realized misconduct risk suggest that such employees may be effective monitors of their firms both vertically (observing upper level management and the control pressures and cultures they create) and horizontally (observing specific acts of misconduct among their colleagues). Similarly, our study complements the small but growing literature on whistleblowers and whistleblowing programs. Whistleblower complaints are typically backward-looking, happening when specific acts of misconduct have already occurred, and whistleblowers often bear considerable personal costs. Both of these features are likely to lead to a "bottleneck" in information transmission about firm risk, with only a fraction of employees who have knowledge of wrongdoing coming forward and likely doing so once risk has been realized. Our results suggest that more widespread information about firm risks may be available via platforms where a broader base of a firm's employees are sharing information that is both relevant and timely as indicators of misconduct risk.

Appendix: Violation Type and Visibility

no.	Visibility	Violation Type	Agency
1	Low	Accounting fraud or deficiencies	Securities and Exchange Commission; U.S. Attorney; Justice Department; Commodity Futures Trading Commission.
2	Low	Agribusiness violation	Grain Inspection, Packers & Stockyards Administration
3	Low	Americans with Disabilities Act	Justice Department
4	Low	Anti-money-laundering deficiencies	Treasury Department Financial Crimes Enforcement Network,U.S. Attorney; Justice Department; Federal Reserve; Securities and Exchange Commission.
5	High	Aviation consumer protection violation	Transportation Department Aviation Consumer Protection Division
6	High	Aviation safety violation	Federal Aviation Administration; U.S. Attorney.
7	Low	Banking violation	Federal Aviation Administration; Justice Department; Federal Deposit Insurance Corporation; Federal Reserve; Office of the Comptroller of the Currency; Consumer Financial Protection Bureau; U.S. Attorney.
8	Low	Benefit plan administrator violation	Employee Benefits Security Administration
8	High	Child labor or youth employment violation	Labor Department Wage and Hour Division
10	Low	Civil contempt	Justice Department
11	High	Consumer protection violation	Federal Trade Commission; Consumer Financial Protection Bureau; Consumer Product Safety Commission; U.S. Attorney.
12	Low	Controlled Substances Act violation	Justice Department; U.S. Attorney; Drug Enforcement Administration.
13	Low	Data submission deficiencies	Commodity Futures Trading Commission; Securities and Exchange Commission.
14	Low	Discriminatory practices	Justice Department; Housing and Urban Development Department; Transportation Department Aviation Consumer Protection Division; U.S. Attorney.
15	High	Drug or medical equipment safety violation	Justice Department; U.S. Attorney; Food and Drug Administration.
16	Low	Economic sanction violation	Office of Foreign Assets Control; Justice Department; Office of Foreign Assets Controls.
17	High	Employment discrimination	Office of Federal Contract Compliance Programs; Equal Employment Opportunity Commission; Justice Department.
18	Low	Energy conservation violation	Energy Department Office of General Counsel
19	Low	Energy market manipulation	Commodity Futures Trading Commission; Federal Energy Regulatory Commission.
20	Low	Energy market violation	Federal Energy Regulatory Commission
21	Low	Environmental violation	Pipeline and Hazardous Materials Safety Administration; Environmental Protection Agency; U.S. Attorney; Bureau of Safety and Environmental Enforcement; Justice Department ; Environmental Protection Agency.
22	Low	Excise tax violation	Alcohol and Tobacco Tax and Trade Bureau
23	Low	Export control violation	Bureau of Industry and Security; State Department Directorate of Defense Trade Controls; U.S. Attorney; Justice Department; National Security Division; Office of Foreign Assets Control.
24	Low	False claims act	Justice Department; U.S. Attorney; Food and Drug Administration.
25	Low	Family and Medical Leave Act	Labor Department Wage and Hour Division

Appendix (cont.)

no.	Visibility	Violation Type	Agency
26	Low	Federal leasing royalty violation	Interior Department Office of Natural Resources Revenue; U.S. Attorney
27	Low	Financial institution supervision failures	Securities and Exchange Commission; Commodity Futures Trading Commission.
28	High	Food safety violation	Food and Drug Administration; Justice Department
29	Low	Foreign corrupt practices act	Securities and Exchange Commission; Justice Department.
30	Low	Foreign exchange market manipulation	Justice Department; Federal Reserve; Commodity Futures Trading Commission.
31	Low	Fraud	U.S. Attorney; Justice Department.
32	Low	Fuel economy (CAFE) violation	National Highway Traffic Safety Administration
33	Low	HHS civil monetary penalties	Health & Human Services Department Office of Inspector General
34	Low	Illicit political contributions	Securities and Exchange Commission
35	Low	Interest rate benchmark manipulation	Commodity Futures Trading Commission; Justice Department; Federal Reserve.
36	Low	Investor protection violation	Securities and Exchange Commission
37	Low	Kickbacks and bribery	U.S. Attorney; Housing and Urban Development Department; Justice Department; Securities and Exchange Commission.
38	High	Labor relations violation	National Labor Relations Board
39	Low	Maritime violation	Federal Maritime Commission
40	Low	Medicare Program violation	Centers for Medicare & Medicaid Services
41	Low	Medicare Parts C and D Enforcement Action	Centers for Medicare & Medicaid Services
42	Low	Mining violation	U.S. Attorney
43	Low	Mortgage abuses	Federal Trade Commission; Justice Department; Housing and Urban Development Department; U.S. Attorney; Federal Reserve; Consumer Financial Protection Bureau
44	High	Motor vehicle safety violation	Federal Motor Carrier Safety Administration; National Highway Traffic Safety Administration; Justice Department.
45	High	Nuclear safety violation	Nuclear Regulatory Commission; Energy Department Office of Enforcement.
46	Low	Nursing home violation	Centers for Medicare & Medicaid Services
47	Low	Off-label or unapproved promotion of medical products	Justice Department; Food and Drug Administration; U.S. Attorney.
48	Low	Offshore drilling violation	Bureau of Safety and Environmental Enforcement
49	Low	Payday lending violation	Consumer Financial Protection Bureau
50	High	Pipeline safety violation	Pipeline and Hazardous Materials Safety Administration
51	Low	Premerger notification violation	Justice Department
52	Low	Price-fixing or anti-competitive practices	Justice Department; Securities and Exchange Commission; U.S. Attorney.
53	Low	Privacy violation	Federal Trade Commission; U.S. Attorney.
54	High	Product safety violation	Consumer Product Safety Commission
55	High	Railroad safety violation	Federal Railroad Administration
56	Low	Securities issuance or trading violation	Commodity Futures Trading Commission; U.S. Attorney; Securities and Exchange Commission.

Appendix (cont.)

no.	Visibility	Violation Type	Agency
57	Low	Service members civil relief act	Justice Department; Office of the Comptroller of the Currency.
58	Low	Student loan abuses	Justice Department; Consumer Financial Protection Bureau.
59	Low	Tax violations	Justice Department; U.S. Attorney; Securities and Exchange Commission.
60	Low	Telecommunications violation	Federal Communications Commission; U.S. Attorney.
61	Low	Tobacco litigation	Justice Department
62	Low	Toxic securities abuses	Securities and Exchange Commission; Justice Department; Fannie Mae; Freddie Mac; National Credit Union Administration; Federal Housing Finance Agency; Federal Deposit Insurance Corporation; U.S. Attorney.
63	High	Uniformed Services Employment and Reemployment Rights Act	Justice Department.
64	High	Wage and hour violation	Labor Department Wage and Hour Division; Labor Commissioner's Office; Department of Workplace Standards; Department of Labor & Industry; Attorney General's Office.
65	Low	Work visa violations	U.S. Attorney; Justice Department.
66	High	Workplace safety or health violation	Occupational Safety & Health Administration; Mine Safety & Health Administration; U.S. Attorney.
67	High	Workplace whistleblower retaliation	Occupational Safety & Health Administration

References

Baucus, M. S. and J. P. Near. 1991. Can illegal corporate behavior be predicted? An event history analysis. *Academy of Management Journal* 34(1): 9–36.

Bloom, N., and J. Van Reenen. 2007. Measuring and explaining management practices across firms and countries. *The quarterly journal of Economics* 122(4), 1351–1408.

Brazel, J. F., K. L. Jones, and M. F. Zimbelman. 2009. Using nonfinancial measures to assess fraud risk. *Journal of Accounting Research* 47(5): 1135–1166.

Caskey, J., Ozel, N. B., 2017. Earnings expectations and employee safety. *Journal of Accounting and Economics* 63(1): 121–141.

Chatman, J. A. and K. A. Jehn. 1994. Assessing the Relationship between Industry Characteristics and Organizational Culture: How Different Can You Be? *The Academy of Management Journal* 37(3): 522–553.

Chen, C. X. and T. Sandino. 2012. Can wages buy honesty? The relationship between relative wages and employee theft. *Journal of Accounting Research* 50(4): 967–1000.

Dechow, P. M., W. Ge, C. R. Larson, and R. G. Sloan. 2011. Predicting material accounting misstatements. *Contemporary accounting research* 28(1): 17–82.

Dyck, A., A. Morse, and L. Zingales. 2010. Who blows the whistle on corporate fraud? *The Journal of Finance* 65(6): 2213–2253.

Finney, H. C., and H. R. Lesieur. 1982. A contingency theory of organizational crime. In S. B. Bacharach (Ed.), Research in the sociology of organizations, vol. 1: 255–299. Greenwich, CT: JAI Press.

Gentzkow, M., B. Kelly, and M. Taddy. 2019. Text as data. *Journal of Economic Literature* 57(3): 535–74.

Good Jobs First. 2019. Accessed 24 Nov 2019. Available at: [www.goodjobsfirst.org/violation–tracker].

Glassdoor. 2019. Accessed 24 Nov 2019. Available at: [https://www.glassdoor.com/index.htm].

Green, T. C., R. Huang, Q. Wen, and D. Zhou. 2017. *Wisdom of the Employee Crowd: Employer Reviews and Stock Return*s. Available at https:/papers.ssrn.com/abstract=3002707.

Hales, J., J. R. Moon, and L. Swenson. 2017. A New Era of Voluntary Disclosure? Empirical Evidence on How Employee Postings on Social Media Relate to Future Corporate Disclosures. *Accounting, Organizations and Society* 68: 88–108.

Heese, J. and G. Perez–Cavazos. When the boss comes to town: the effects of headquarters' visits on facility–level misconduct. *The Accounting Review*: In Press.

Henry, E. and A. J. Leone. 2015. Measuring qualitative information in capital markets research: Comparison of alternative methodologies to measure disclosure tone. *The Accounting Review* 91(1): 153–178.

Hobson, J. L., W. J. Mayew, and M. Venkatachalam. 2012. Analyzing speech to detect financial misreporting. *Journal of Accounting Research* 50(2): 349–392.

Huang, X., S. H. Teoh, and Y. Zhang. 2013. Tone management. *The Accounting Review* 89(3): 1083–1113.

Huang, X., A. Nekrasov, S. H. Teoh. 2018. Headline Salience, Managerial Opportunism, and Over– and Underreaction to Earnings. *The Accounting Review*. In–Press.

Jia, Y., L. V. LENT, and Y. Zeng. 2014. Masculinity, testosterone, and financial misreporting. *Journal of Accounting Research* 52(5): 1195–1246.

Larcker, D.F., S. A. Richardson, and I. Tuna. 2007. Corporate Governance, Accounting Outcomes, and Organizational Performance. *The Accounting Review* 82(4): 963–1008.

Liu, X. 2016. Corruption culture and corporate misconduct. *Journal of Financial Economics* 122(2):307–327.

Loper, E., and S. Bird. 2002. NLTK: the natural language toolkit. *arXiv preprint cs/0205028*.

Maas, V. S. and M. Van Rinsum, 2013. How control system design influences performance misreporting. *Journal of Accounting Research* 51(5): 1159–1186.

O'Reilly, C. A., J. Chatman, and D. F. Caldwell. 1991. People and Organizational Culture: A Profile Comparison Approach to Assessing Person–Organization Fit. *The Academy of Management Journal* 34(3): 487–516.

Porter, M. F. 1980. An algorithm for suffix stripping. *Program* 14(3): 130–137.

Purda, L., and D. Skillicorn. 2015. Accounting variables, deception, and a bag of words: Assessing the tools of fraud detection. *Contemporary Accounting Research* 32(3): 1193–1223.

Rapp, G. C. 2010. False Claims, Not Securities Fraud: Towards Corporate Governance by Whistleblowers. *Nexus Journal of Law & Policy* 55: 101–111.

———. 2012. States of Pay: Emerging Trends in State Whistleblower Bounty Schemes. *South Texas Law Review* 54: 53–79.

Rogers, J. L., A. Van Buskirk, and S. L.Zechman. 2011. Disclosure tone and shareholder litigation. *The Accounting Review* 86(6): 2155–2183.

Soltes, E. 2019. The frequency of corporate misconduct: public enforcement versus private reality. *Journal of Financial Crime* 26 (4): 923–937.

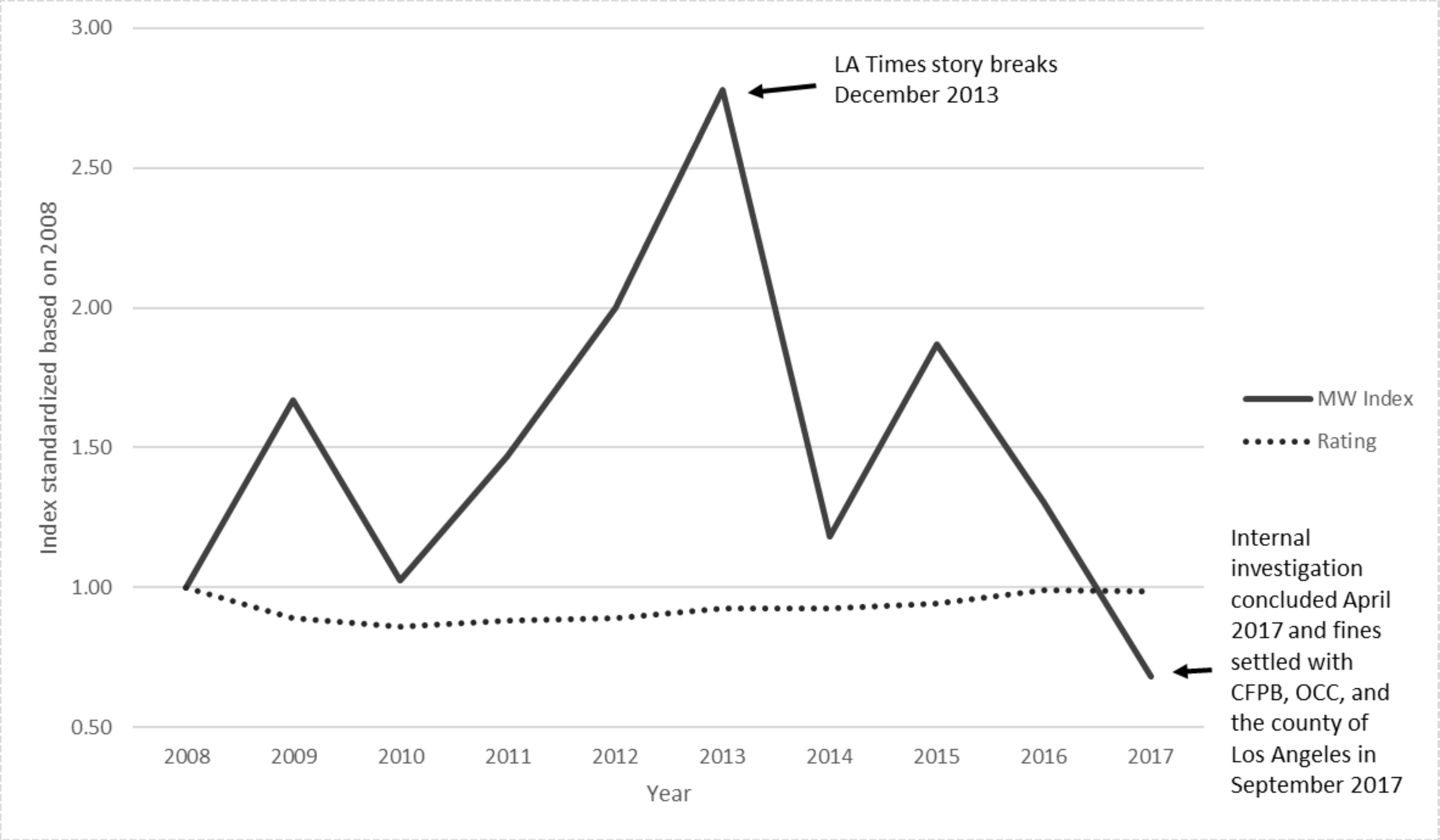
Srinivasan, S, D.W. Campbell, S. Gallani, and A. Migdal. 2017. "Sales Misconduct at Wells Fargo Community Bank." *Harvard Business School Case* 118–009.

Taddy, M. 2013. Multinomial inverse regression for text analysis. *Journal of the American Statistical Association* 108(503): 755–770.

Uzun, H., S. H. Szewczyk, and R. Varma. 2004. Board composition and corporate fraud. *Financial Analysts Journal* 60(3): 33–43.

Vaughan, D. 1982. Toward understanding unlawful organizational behavior. *Michigan Law Review* 80: 1377–1402.

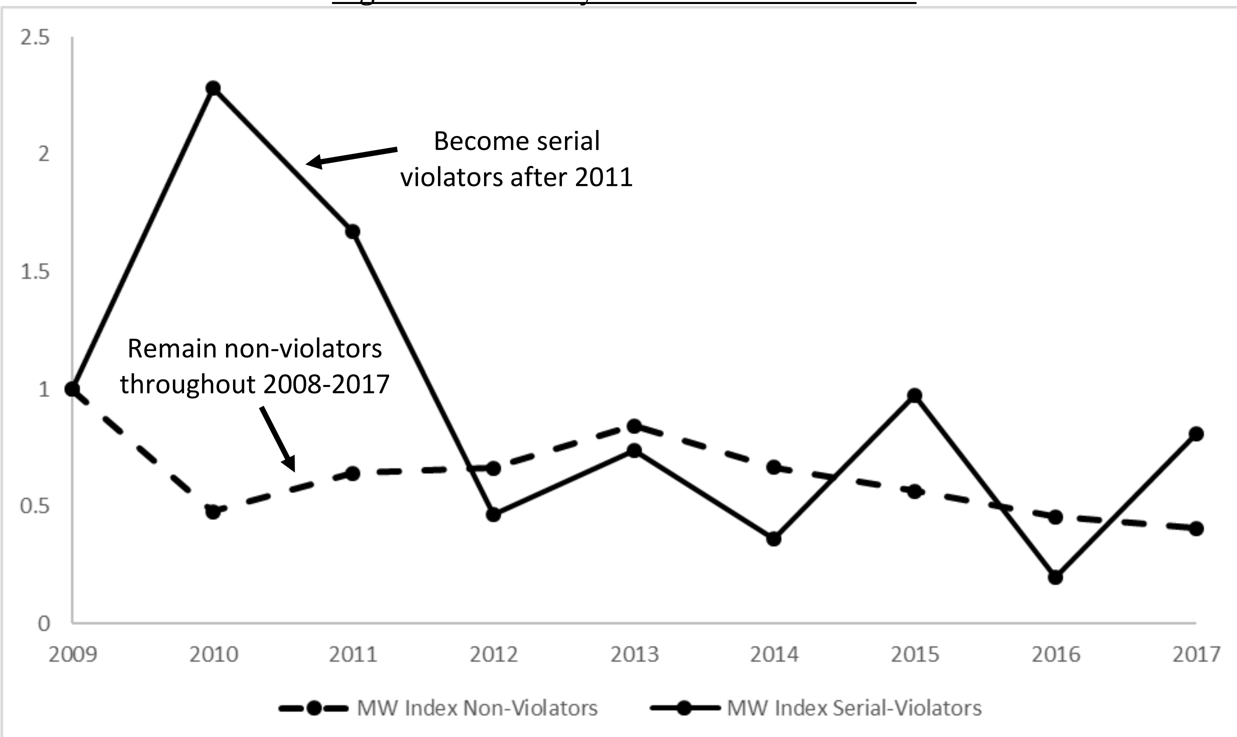
Figure 1: Misconduct Word Index and Glassdoor Ratings during a Period of Known Widespread Misconduct in Sales Practices at Wells Fargo



Note: Figure 1 shows a summary misconduct word index ("MW Index") as well as average Glassdoor ratings ("Rating") for Wells Fargo over the period 2008-2017. The summary misconduct word index is the sum of our various misconduct word indices. For comparability, both the summary index and Glassdoor ratings are normalized to 1 in 2008.

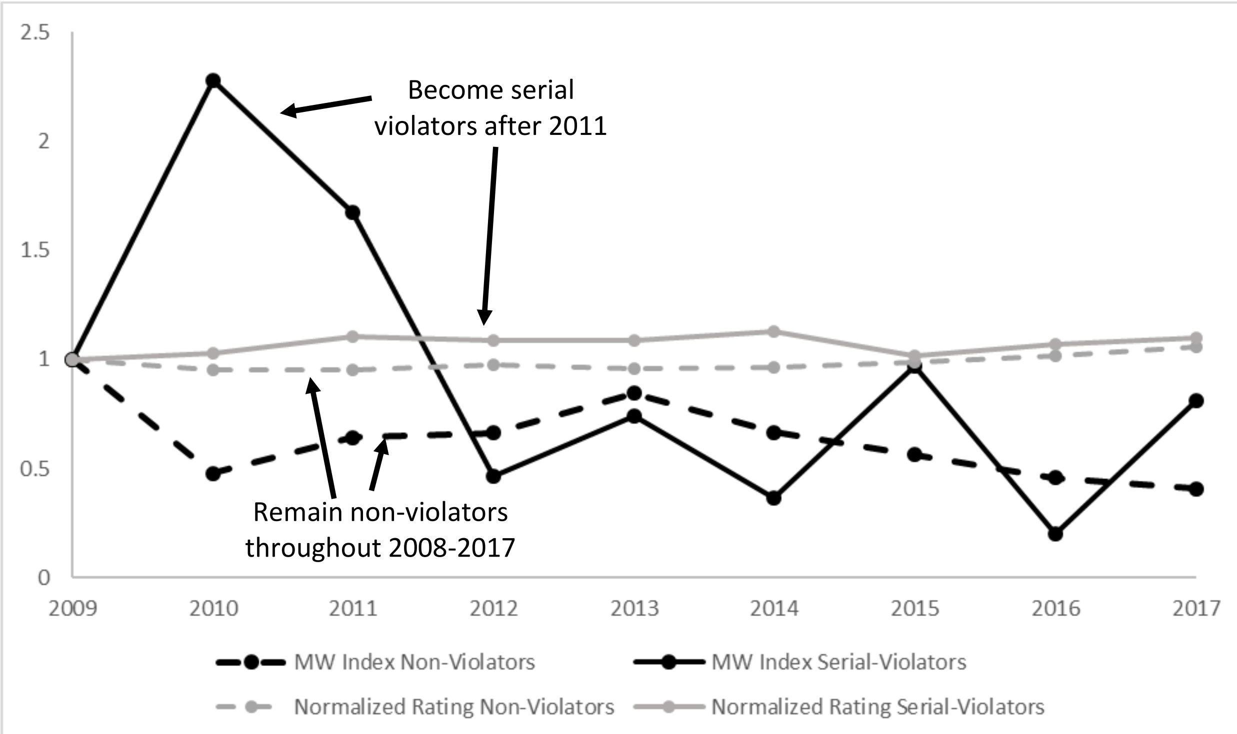
Figure 2: Misconduct Word Index for Firms that Transitioned from Non-Violator Status (2008-2012) to Serial Violator Status (2012-2017)

Figure 2a: Summary Misconduct Word Index



Note: Figure 2a shows a summary misconduct word index ("MW Index") for two subsamples of firms, both of which had no violations over the period 2008-2011. "Non-violators" are those firms that also had no violations during 2012-2017 (N=654 firms). "Serial violators" are those firms that went on to have violations in at least two of the years during the period 2012-2017 (N=126 firms). The summary misconduct word index is the sum of our various misconduct word indices. For comparability, the summary indices are normalized to 1 in 2009.

Figure 2b: Including Normalized Ratings



Note: Figure 2b shows a summary misconduct word index ("MW Index") as well as average Glassdoor ratings ("Rating") for two subsamples of firms, both of which had no violations over the period 2008-2011. "Non-violators" are those firms that also had no violations during 2012-2017 (N=654). "Serial violators" are those firms that went on to have violations in at least two of the years in the period 2012-2017 (N=126). The summary misconduct word index is the sum of our various misconduct word indices. For comparability, both the summary index and Glassdoor ratings are normalized to 1 in 2009.

Table 1: Descriptive Statistics

Variable	Description	Mean	Standard Deviation	10th Percentile	Median	90th Percentile
<u>OutcomeVariables</u>						
ViolationHV	Indicator for the presence of any “high-visibility” violation directly related to workers or consumers (e.g. safety, health, employment, wages, hours, and labor relations; product/service safety, consumer protection, etc...)	0.178	0.382	0	0	1
ViolationLV	Indicator for the presence of any “low-visibility” violation more likely to occur at the management level of the firm (e.g. political contributions, bribery, kickbacks, money laundering; tax fraud; price fixing; accounting fraud; insider trading, etc…)	0.093	0.29	0	0	0
#ViolationsHV	Total number of “high-visibility” violations directly related to workers or consumers (e.g. safety, health, employment, wages, hours, and labor relations; product/service safety, consumer protection, etc...)	45.30	1,204	0	0	2
#ViolationsLV	Total number of “low-visibility” violations more likely to occur at the management level of the firm (e.g. political contributions, bribery, kickbacks, money laundering; tax fraud; price fixing; accounting fraud; insider trading, etc…)	1.53	19.51	0	0	0
Penalty	Penalties imposed by the relevant regulatory or legal authorities	$3,865,126	$178,000,000	0	0	$5,600
<u>Glassdoor Data</u>						
Rating	Average overall Glassdoor rating	2.73	0.78	2	3	4
Number of Reviews	Number of employee reviews on Glassdoor	61.01	241.73	2	10	171
Number of Words	Total number of words across Glassdoor employee reviews	2,003	6,832	56	381	5,588
<u>Firm Characteristics</u>						
Size	Natural log of assets	7.82	2.14	5.1	7.78	10.47
Leverage	Debt/Equity	0.78	63.46	0	0.474	2.48
ROA	End of year return on assets	-0.006	1.41	-0.11	0.031	0.12

Note: N=10,156 firm-year observations representing 1,388 firms over 10 years from 2008-2017

Table 2: Top 25 Weighted Words in Misconduct Word Indices

Any Violation (High Visibility)			Any Violation (Low Visibility)			Serial Violator		
Pros	Cons	Advice	Pros	Cons	Advice	Pros	Cons	Advice
benefit	Zero	great	opportunity	Posit	employee	manage	work	manage
company	Level	worker	within	Downside	none	Work	approve	improve
employee	Staff	younger	lot	Place	great	Easy	promotion	know
train	Little	listen	vacation	Salary	leadership	Well	keep	fair
something	require	hone	company	Thing	clerk	Make	influence	culture
make	Zone	Time	grow	Progress	realize	Culture	process	sale
opportunity	manage	Fair	conservative	Compensation	show	Change	kind	make
inform	engine	people	work	Take	integration	Loyalty	strange	environment
move	sometime	Start	plan	Product	value	Issue	review	well
hard	business	Like	pretty	Lack	review	flexible	support	vision
structure	downside	experience	area	Associate	communicate	interact	respect	profit
quality	Need	strategic	allow	Success	lead	Short	lack	money
safety	number	integration	contact	Close	appreciation	Care	understand	structure
worker	treatment	outside	strong	Abandon	zone	communication	leadership	tough
promote	Boss	ahead	best	Get	come	Leader	stress	push
pretty	Think	store	cash	Look	back	Allow	balance	overwork
accept	everything	Zero	challenging	Mental	greater	Sale	report	learn
advance	Raise	standard	hard	Make	job	Stay	issue	hire
employ	smaller	safety	sale	Poor	within	process	horrible	motivation
feel	Able	business	site	Security	level	Able	help	allow
corpora	Hire	office	integrity	Home	many	Trouble	detail	career
help	supervision	offer	project	Lunch	sale	Hard	difficult	appreciation
flexible	difficult	product	exposure	Engine	philosophy	generous	lie	analyze
bigger	associate	Help	move	Term	product	Hire	analysis	leadership
really	Good	Live	product	Downtown	idea	overtime	target	force

Note: Table shows the top 25 words in the pros, cons, and advice sections of Glassdoor ratings which receive the highest weights in calculating our misconduct word indices. Columns 1-3 (4-6) are the words that load most highly for the presence of any high (low) visibility violation in a given firm-year. Columns 7-9 are the words that load most highly for serial violating firms, defined as firms that had violations in at least 2 years during the period 2008-2011.

Table 3: Correlations between Text Measures and Ratings

	Rating	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*
(1) *MW_Index_HV – Pros*	-0.094*					
(2) *MW_Index_HV - Cons*	-0.106*	0.060*				
(3) *MW_Index_HV - Advice*	-0.002	0.033*	0.064*			
(4) *MW_Index_LV - Pros*	0.070*	0.215*	-0.009	0.026*		
(5) *MW_Index_LV - Cons*	-0.087*	-0.014	-0.278*	-0.035*	-0.006	
(6) *MW_Index_LV - Advice*	-0.098*	-0.017	0.008	-0.541*	-0.035*	-0.014

Note: Table shows correlations between weighted misconduct word indices and overall ratings. *MW_Index_HV* and *MW_Index_LV* are the misconduct word indices for high and low visibility violations respectively and are calculated separately for the pros, cons, and advice sections respectively. * Denotes significance at the p<0.10 level.

Table 4: Performance and Variable Influence for Violation Prediction Models

	Prediction Performance in Test Sample		*Influence Statistics*							
Outcome	*R-Squared*	*% Correct Predictions*	*Lagged Outcome*	*MW Index - Pros*	*MW Index- Cons*	*MW Index - Advice*	*Rating*	*Size*	*Leverage*	*ROA*
Violation - High Visibility	27.9%	93.9%	34.6%	4.9%	3.5%	5.6%	1.0%	23.2%	5.9%	5.8%
Violation - Low Visibility	26.7%	94.0%	23.1%	5.6%	4.5%	4.3%	1.3%	21.9%	9.0%	10.2%
Violation - Workplace Safety	22.2%	93.2%	39.5%	2.8%	3.9%	3.0%	0.4%	19.4%	5.1%	5.6%

Note: Table shows results from fitting a logistic model using a gradient boosted tree, where the dependent variable is an indicator for the occurrence of a violation during year 't'. The predictors include firm characteristics (*Size, Leverage, ROA,* and an indicator for the occurrence of a violation in the prior year); indicators for industry and year; and data from Glassdoor (average overall rating and our misconduct word indices). The test sample consists of 20% of firms held out from the training set and is used to assess prediction performance of the final model. Prediction performance is measured using the pseudo-R-squared, calculated as R2=1-L1/L0 where L1 and L0 are the log likelihood of the full model and intercept-only models, respectively. Influence statistics measure the relative contribution of each variable to the improvement in the log likelihood across training iterations. *Correct p(Outcome)* is the proportion of correct predictions from the final model in the test sample.

Table 5: Prediction Performance and Variable Influence for Different Violation Outcomes

	Prediction Performance in Test Sample	*Influence Statistics*							
Outcome	*R-Squared*	*Lagged Outcome*	*SR - Pros*	*SR - Cons*	*SR - Advice*	*Rating*	*Size*	*Leverage*	*ROA*
#Violations	38.5%	68.0%	1.9%	4.2%	5.7%	2.2%	0.4%	2.9%	0.2%
MultipleViolations	58.0%	32.8%	4.1%	10.9%	6.1%	2.0%	10.5%	4.7%	4.3%
Penalty	31.9%	33.4%	8.5%	5.5%	4.8%	0.3%	13.2%	6.2%	5.7%

Note: Table shows results from fitting gradient boosted trees, where the dependent variables are the total number of violations (*#Violations*); a categorical variable indicating 0, 1, and more than 1 violations respectively (*MultipleViolations*); and the natural log of (1+*Penalty*). The predictors include firm characteristics (*Size, Leverage, ROA*, and a one year lag of the outcome of interest); indicators for industry and year; and data from Glassdoor (average overall rating and our misconduct word indices). The test sample consists of 20% of firms held out from the training set and is used to assess prediction performance of the final models. Gradient boosting is used to fit poisson, multinomial, and normal regression models for predicting *#Violations, MultipleViolations,* and ln(*1+Penalty*), respectively. Prediction performance is measured using either traditional R-squared (in the case of *ln(1+Penalty)*) or pseudo-R-squared (in the cases of *#Violations* and *MultipleViolations*). Pseudo-R-squared is calculated as R2=1-L1/L0 where L1 and L0 are the log likelihood of the full model and intercept-only models, respectively. Influence statistics measure the relative contribution of each variable to the improvement in the log likelihood across training iterations.

Table 6: Predictive Performance and Variable Influence for Serial Violator Transition Models

	Prediction Performance in Test Sample		*Influence Statistics*							
Outcome	*R-Squared*	*% Correct Predictions*	*Any Prior Violation*	*MW Index-Pros*	*MW Index-Cons*	*MW Index-Advice*	*Rating*	*Size*	*Leverage*	*ROA*
Serial Violator - High Visibility	35.2%	91.1%	12.5%	10.3%	9.2%	10.8%	3.6%	9.3%	8.8%	7.1%
Serial Violator - Low Visibility	10.5%	90.1%	3.7%	9.7%	8.0%	9.4%	4.9%	24.7%	5.3%	6.3%

Note: Table shows results from fitting a logistic model using a gradient boosted tree, where the dependent variable is an indicator for whether a non-serial violating firm from 2008-2011 becomes a serial violator between 2012 and 2017. We define a serial violator as a firm that has 2 or more violation years during a given period and a non-serial violator as a firm that has 1 or 0 years with violations during that period. The predictors include firm characteristics (*Size*, *Leverage*, *ROA*, industry indicators, and an indicator for the occurrence of any violation in the period 2008-2011); and data from Glassdoor (average overall rating and our misconduct word indices). The test sample consists of 20% of firms held out from the training set and is used to assess prediction performance of the final model. Prediction performance is measured using the pseudo-R-squared, calculated as R2=1-L1/L0 where L1 and L0 are the log likelihood of the full model and intercept-only models, respectively. Influence statistics measure the relative contribution of each variable to the improvement in the log likelihood across training iterations. *% Correct Predictions* is the proportion of correct predictions from the final model in the test sample.