Are Firms Sacrificing Flexibility for Diversity and Inclusion?

Hoa Briscoe-Tran*

August 30, 2024

Abstract

This paper examines how workforce diversity and inclusion (D&I) affects corporate flexibility – a firm's ability to adapt to changes, crucial for success and survival during crises. Economic theories suggest D&I may enhance problem-solving but increase decision-making delays. We explore these trade-offs by extrapolating a novel 2020 employee D&I rating back to 2008 for thousands of companies using a supervised large language model. Analyses show that D&I firms exhibit lower flexibility, with evidence suggesting a causal relationship. Further investigation of firms' responses to a major economic shock reveals that D&I is associated with adjustment costs in operating efficiency rather than workforce management.

Keywords: Flexibility, D&I, DEI, operating flexibility, firm performance, workplace diversity, NLP, BERT, employee ratings, organizational behavior, causal effects, human capital.
JEL Classification: G30, G34, M14, D22, J71, L25, D63, J24.

* Alberta School of Business, the University of Alberta. This paper has benefited from comments by Jocelyn Evans, Isil Erel, Laura Field, Simon Glossner, John Graham, Isaac Hacamo, Davidson Heath, Susan Jackson, Ralph Koijen, Tong Li, Michelle Lowry (discussant), Xinjie Ma, Yueran Ma, Hernan Ortiz-Molina, Edmund Phelps, Amin Shams, René Stulz, Sorabh Tomar, Rohan Williamson, Michael Wittry, Alexander Zentefis, seminar participants from Ohio State University and University of Alberta, and the mentors at the Rising Scholars Conference at Chicago Booth, the Diversity Emerging Scholars Program, the corresponding Best Papers Session, and the University of Connecticut Finance Conference. We thank Yue Qiu for sharing the H1-B labor data.

1. Introduction

Investors, policymakers, and the general public have increasingly focused on the diversity and inclusion (D&I) practices of various organizations, from state employers to private sector corporations. This shift extends beyond the top management or board level to the broader workforce, especially following influential social movements like Me Too and Black Lives Matter. While implementing effective D&I practices aligns with social justice considerations, it is unclear whether and how these practices affect an organization's operations and performance.

Take profit-maximizing firms, for example. If D&I practices unequivocally enhanced all dimensions of a firm's performance, we would expect to see universal and increasing adoption. Of course, that is not the case, so there must be an economic trade-off with D&I practices.

A thorough review of economic theories reveals potential economic costs associated with D&I practices, independent of recent political backlash against these initiatives across the US. Diversity may create communication challenges (Lang (1986)), increase miscoordination (e.g., Milliken and Martins (1996), Alesina, Baqir, and Easterly (1999), Kets and Sandroni (2021)), and reduce trust (Alesina and La Ferrara (2002)). Furthermore, fostering inclusiveness among diverse agents often requires time (e.g., Jackson (1992)), which could slow decision-making and implementation processes. Additionally, D&I practices targeting hiring, firing, and promotion decisions may constrain an organization's workforce management (Athey, Avery, and Zemsky (2000), Fryer Jr and Loury (2013)), essentially increasing labor adjustment costs. Collectively, these factors suggest that diverse and inclusive organizations may exhibit lower flexibility, i.e., lower ability to adapt effectively to economic changes.

Nonetheless, there are reasons to expect the opposite as well, that D&I could help an organization become more flexible in its operations. A diverse workforce could provide an organization with more information and approaches to problem-solving (Mannix and Neale (2005)), enabling better decisions in adapting to changes. Moreover, D&I practices can help an organization gather more customer support (Hacamo (2023)) and attract more prospective employees (Avery and McKay (2006), Choi et al. (2023)), allowing for more flexible expansion when needed. So overall, it is an empirical question how D&I affects an organization's flexibility.

Our focus on organizational flexibility not only evaluates theories about D&I's economic trade-offs but also addresses a critical aspect of organizational success. Barry, Campello, Graham and Ma (2022) highlight the importance of corporate flexibility for firms' survival and performance

during crises. While previous research has primarily examined team-level diversity or diversity without inclusion, our study contributes by investigating firm-level D&I effects, which may not directly translate from team-level findings (e.g., Klein and Kozlowski (2000)). We develop a novel, comprehensive measure of organizational D&I and link it to flexibility, utilizing the rich accounting data available for publicly listed firms. This approach fills a significant gap, as no prior studies have examined D&I's impact on organizational flexibility.

Specifically, we utilize a novel employee rating of D&I from Glassdoor.com, a widely used career intelligence website. Leveraging Glassdoor's pioneering D&I rating introduced in 2020, we extrapolate the rating back to 2008 based on review texts, via a cornerstone large language model from Google, BERT. After demonstrating the reliability of the extrapolated rating as a measure of D&I, we use it to examine the relationship between a firm's D&I rating and its flexibility, based on the flexibility measure that Gu, Hackbarth, and Johnson (2018) derive from accounting data. Higher D&I ratings correlate with lower flexibility, with some causal evidence from an exogenous D&I shock. Analysis of D&I firms during a major economic crisis confirms reduced flexibility due to larger operating efficiency declines, suggesting D&I-related decision-making frictions.

Establishing these results requires properly defining and measuring diversity and inclusion (D&I). According to Harvard Business Publishing, *diversity* encompasses human demographic differences and the variety of ideas, backgrounds, and opinions people bring; whereas *inclusion* refers to employees being valued, respected, and encouraged to fully participate. These two concepts are often defined together because it is increasingly recognized that diversity without inclusion is insufficient to influence workplace culture and business outcomes (Nishii (2013)).

The above definition of D&I indicates three major measurement challenges. First, a D&I measure must capture multiple dimensions, ranging from more visible attributes like race and gender to less visible attributes like sexual orientation and life experiences. Second, even measuring D&I only on race and gender proves difficult because most companies do not publicly disclose this information.[1] Third, and perhaps most importantly, a D&I measure must capture inclusion beyond diversity, which refers to whether employees are valued and encouraged to participate, regardless of their differences. Therefore, simply counting minorities in a firm's workforce is not enough to capture D&I.

[1] 68% of Russell 1000 firms have no disclosure on workforce race/gender in 2021. See https://justcapital.com/news/a-small-fraction-of-corporations-share-diversity-data-but-disclosure-is-rapidly-on-the-rise/.

Using employee reviews can address all the three challenges of measuring D&I. First, in reviews, employees often write anonymously about working at a firm and often mention multiple dimensions of the workplace, including both more visible attributes like racial diversity and less visible attributes like LGBT friendliness. Second, employee reviews data have become increasingly abundant, especially since 2008, covering at least 500,000 US employers by 2021. The rich data, thus, enable measuring D&I for a comprehensive sample of companies over many years. Third, employee reviews capture employees' perception of how they are treated, so it likely captures inclusion beyond diversity.

We specifically examine employee reviews from Glassdoor for two reasons. First, it is a career intelligence site well-known for its balanced and informative reviews (Marinescu et al. (2021), Green et al. (2019)). Second, and most importantly, Glassdoor pioneered a diversity and inclusion rating in 2020 that we can leverage to create a consistent D&I measure across firms over many years. The new feature allows employees to anonymously rate a firm's D&I practices on a five-star scale.

Exploiting the newly introduced D&I rating, we develop a model to estimate the D&I rating for every review since Glassdoor's inception in 2008, based on the review's other ratings and written texts. Our model can understand review texts well because it utilizes BERT, a language model by Google that outperforms humans on reading comprehension (Devlin et al. (2018)). BERT's pre-training on a vast corpus allows it to transfer its extensive knowledge of the English language to the context of employee reviews. By fine-tuning the BERT model to estimate the D&I rating for 10.4 million reviews in our sample, we obtain the average D&I rating for each firm-year, serving as our firm-level D&I measure.

Before examining the firm-level measure, we assess the estimated D&I rating at the review level, in two ways. First, we follow Acikalin et al. (2022) to identify the words most commonly used by our model and find them to be very much D&I-related, featuring *inclusive*, *diverse*, and *mentorship* in positive contexts, and *racist*, *discrimination*, and *harassment* in negative contexts. These words also suggest that our model does capture D&I, with an emphasis on inclusion beyond diversity. Second, we follow a dictionary approach to directly target D&I words in employee reviews, creating a measure that validates our model. Interestingly, the D&I rating predicted by our model exhibits a higher correlation with the dictionary-based measure than the actual D&I rating provided by employees.

The D&I rating at the firm-year level also appears reliable. First, the measure strongly predicts the future likelihood of a firm appearing on Fortune magazine's 100 Best Workplaces for Diversity, a list constructed from surveying minorities within a firm, such as women, people of color, and LGBTQ people. Second, the D&I measure's predictive power remains strong after controlling for a firm's D&I policies and also is unaffected by a look-ahead bias (see Section 3). Third, because D&I is an important component of corporate culture (Guiso, Sapienza, and Zingales (2015, p. 64)) and is likely the culture value most relevant to employees, a good D&I measure should be highly correlated with how employees view corporate culture. We find that our D&I measure has a strong correlation (0.90) with employees' rating of culture on Glassdoor. Nonetheless, our D&I measure still significantly predicts the Best Diversity list after controlling for the culture rating. Finally, we show that the lagged value of our D&I measure significantly predicts itself beyond all the other ratings on Glassdoor, suggesting that the D&I measure captures unique information beyond the other Glassdoor ratings.

Using the D&I measure, we next examine the relationship between D&I and a firm's flexibility, defined as its ability to adjust and adapt to changes in economic conditions (Barry et al. (2022)). Following Gu, Hackbarth, and Johnson (2018), we gauge a firm's flexibility from the range of the firm's historical cost margins. The idea is that a less flexible firm would adapt less to economic shocks, allowing these shocks to bring the firm's operating cost margins to more extreme levels (i.e., a wider range). We find that firms with a higher D&I rating during 2009-2021 exhibit significantly lower flexibility. Compared to a 3-star D&I rated firm, a 4-star rated firm operates on an operating cost range that is 0.36 standard deviation higher. This finding remains significant after controlling for industry fixed effects, governance attributes, diversity indicators at the board and top management levels, as well as *financial flexibility* measures like cash and debt ratios, suggesting that better D&I is associated with lower *operating flexibility*.

While the negative relationship between D&I and flexibility holds after controlling for many factors, it may still be driven by some unobservable factors. To alleviate this concern, we next study how flexibility changes following a plausibly exogenous shock to D&I specifically. We examine a court ruling that unexpectedly increased firms' incentives to improve D&I practices. Prior to 2013, US employers were held liable for workplace harassment only when the harasser held a supervisory role over the victim. In 2013, however, the 7th Circuit Court, which sets legal precedents for cases in Illinois, Indiana, and Wisconsin, unexpectedly ruled against an employer

for racial and sexual harassment even when the harasser was merely a co-worker, not a supervisor, of the victim. This court ruling increased the likelihood of harassment lawsuits, prompting firms in the three states (treated firms) to significantly enhance their D&I practices compared to other firms (control firms), particularly when the firms had poor D&I practices prior to the ruling. We find that to be the case, as D&I ratings increased while flexibility declined following the court ruling for the treated relative to control firms, especially among the firms with a low prior D&I rating. The results imply that D&I improvements lead to a decline in flexibility, i.e., a causal relationship.

Two potential channels may explain why D&I reduces a firm's flexibility. One is the *workforce* channel, where D&I considerations introduce hiring and firing frictions that restrict a firm's ability to manage its workforce. Another is the *efficiency* channel, where D&I firms face communication challenges among diverse groups, such as more conflicts and more time to integrate differences across employees, leading to slower decision-making and hindering agile adaptation to changes.

To explore these channels, we study how D&I firms manage workforce and operating efficiency following a large economic shock. In our sample, the COVID crisis in 2020 provided an ideal and significant unexpected shock, allowing us to observe firms' flexibility in response. We measure a firm's COVID exposure by its operational reliance on in-person interactions (Koren and Peto (2020)), and confirm that D&I firms' operating performance was significantly more affected by their COVID exposure, indicating lower flexibility in adapting to the shock. The relative decline in performance, however, was not due to D&I firms failing to cut their workforce more than other firms during 2020. Instead, we find that D&I firms experienced a larger decline in operating efficiency, proxied by sales per employee, sales growth, and assets turnover. These results remain similar after controlling for firms' cash and debt holdings, governance indicators, board diversity, customer base, labor intensity, labor skill (H1-B workers), and industry trends. Overall, the analyses suggest that D&I firms exhibit lower flexibility because they have more frictions on operating efficiency rather than frictions with workforce management in adapting to changes.

Literature: This paper contributes to several literatures. First, it adds to the literature on corporate flexibility, which has gathered increasing attention in the literature because of the role flexibility played during the recent COVID crisis (Barry et al. (2022)). Previous studies have underscored the importance of corporate flexibility, influencing aspects such as firm risk and stock

returns (Hackbarth and Johnson (2015), Gu, Hackbarth, and Johnson (2018)), cash holdings (Ghaly, Anh Dang, and Stathopoulos (2017)), as well as capital structure (Simintzi, Vig, and Volpin (2015), Serfling (2016), Reinartz and Schmid (2016), D'Acunto et al. (2018), Gu and Hackbarth (2021)). However, the origins of corporate flexibility have remained relatively unexplored. Our paper addresses this gap by highlighting the significant impact of workforce D&I practices on corporate flexibility. In this way, the paper adds to the labor and finance literature as well (Matsa (2018)).

Second, this paper contributes to the literature on corporate social responsibility (CSR) and environmental, social, and governance (ESG) considerations in (Gillan, Koch, and Starks (2021)). Most papers in this literature study the broad question of whether "ESG/CSR activities [are] beneficial to shareholders" by considering ESG/CSR as a whole, or at best, E, S, and G as the big categories. However, different ESG activities could have different, if not opposite, effects on a firm's operations, policies, and performance (Edmans (2023)). In this regard, our paper focuses on diversity and inclusion (D&I), a narrower and more well-defined issue that has recently attracted substantial media attention. D&I is not merely a sub-component of ESG, but a fundamental aspect of business operations with potential far-reaching impacts on team performance (e.g., Lu, Naik, and Teo (2024)), innovation (e.g., Griffin, Li, and Xu (2021)), and risk management (e.g., Bernile, Bhagwat, and Yonker (2018)).

Third, this paper contributes to the literature on discrimination and diversity across many fields. While the economics and finance literature has studied these topics via the difference between minority and majority groups on outcomes, such as hiring outcomes (Bertrand and Duflo (2017), Benson, Board, and Meyer-ter-Vehn (2023)), wages (Rubinstein and Brenner (2014), Lang and Kahn-Lang Spitzer (2020)), capital allocation (Duchin, Simutin, and Sosyura (2021)), and loan approvals (Butler, Mayer, and Weston (2022), Frame et al. (2021)), the management literature has long called for "approaches that take into account the subjective experiences of the group members" (Mannix and Neale (2005, p. 39)). Our paper responds directly to this call by constructing a D&I measure from employees' perception of their treatment.

In addition, while most of the literature focuses on diversity at the board or top management level (e.g., Hambrick, Cho, and Chen (1996), Bernile, Bhagwat, and Yonker (2018)), our paper focuses on diversity and inclusion at the workforce level, which we show to have an effect on corporate flexibility beyond top management or board diversity. Even at the board level with more

available data, most studies focus on diversity alone without inclusion metrics, despite a lack of inclusiveness in boardrooms (Field, Souther, and Yore (2020)). By contrast, this paper studies diversity and inclusion together.

Despite the attention to D&I issues, studies of workforce D&I practices have been limited by a lack of a comprehensive firm-level D&I measure. One exception is Edmans, Flammer, and Glossner (2023), who create a novel D&I measure by selecting D&I-relevant questions from employee surveys in Fortune's Best Companies to Work For competition. Our approach differs by using employees' direct perceptions of D&I practices, which we show to align with a general D&I definition, rather than researchers' selection of relevant survey items. Our study also encompasses a broader sample of over 2000 firms, while they focus on a much smaller sample, perhaps because only a subset of firms participate in the Fortune competition. Lastly, while Edmans et al. study the association between D&I and a firm's accounting and stock performance, this paper studies the effect of D&I on a firm's flexibility, a less studied but important issue, with some causal evidence.

Finally, this paper contributes to the rapidly expanding literature on applications of large language models like BERT in finance, accounting, and economics. Typical applications use BERT to learn from researchers' manual text labeling to capture constructs such as corporate goals (Rajan, Ramella, and Zingales (2023)), climate risk disclosure (Kölbel et al. (2022)), labor shortage (Harford, He, and Qiu (2023)), and inflation exposure (Chava et al. (2022)), among others. As Rajan, Ramella, and Zingales (2023) point out, this reliance on manual classification is time-consuming and generates limited training data. Our approach differs by having BERT learn directly from employees' provided D&I ratings, offering abundant training data and greater replicability.

This paper proceeds as follows. Section 2 defines D&I and corporate flexibility, and develops the hypotheses connecting these concepts. Section 3 describes the data and details how we measure a firm's D&I practices. Section 4 evaluates whether D&I firms exhibit more flexibility than others. Section 5 investigates the mechanisms of how D&I affects corporate flexibility, Section 6 discusses various robustness checks, and Section 7 concludes.

2. Hypothesis development

In this section, we clarify the definitions of diversity and inclusion (D&I) as well as corporate flexibility, and develop hypotheses about how D&I could affect a firm's flexibility.

2.1. Definitions of D&I and flexibility

Diversity and Inclusion (D&I)

The nonprofit Harvard Business Publishing defines diversity and inclusion as follows:[2]

> **"Diversity** refers to anything that sets one individual apart from another, including the full spectrum of human demographic differences as well as the different ideas, backgrounds, and opinions people bring.
>
> **Inclusion** implies a cultural and environmental feeling of belonging and sense of uniqueness. It represents the extent to which employees feel valued, respected, encouraged to fully participate, and able to be their authentic selves."

The above definitions coincide with various definitions from business practitioners and the academic literature on diversity management. Roberson (2006) asks human resource officers in 51 large corporations to define diversity and inclusion. She concludes that "definitions of diversity focused primarily on differences and the demographic composition of groups or organizations, whereas definitions of inclusion focused on organizational objectives designed to increase the participation of all employees and to leverage diversity effects on the organization." (p. 12). So, while the general definitions of D&I do not directly imply business benefits, business practitioners often view D&I as an organizational feature that a firm could leverage for business benefits.

Diversity and inclusion are defined together because of the increasing recognition that "diversity is useless without inclusivity" (Riordan (2014)). Employees from diverse social and cultural groups are often excluded from networks of information and opportunity in a company (Ibarra (1993)). So, while a firm could have employees with a variety of backgrounds and skillsets, the firm might not utilize these diverse human resources to enhance its performance. This perspective is not new (e.g., Prasad (2001)), but has grown into mainstream diversity management recently (e.g., Sherbin and Rashid (2017), Holmes et al. (2021)).

Flexibility

Barry et al. (2022) define corporate flexibility as "the ability of firms to adjust and adapt" to changing economic conditions. While general, this definition encompasses many dimensions of corporate flexibility. One widely studied dimension is financial flexibility, which Denis (2011) defines as "the ability of a firm to respond in a timely and value-maximizing manner to unexpected

[2] https://www.harvardbusiness.org/start-here-a-primer-on-diversity-and-inclusion-part-1-of-2/

changes in the firm's cash flows or investment opportunity set." Other dimensions of corporate flexibility, so-called operating flexibility, includes workforce flexibility (Simintzi, Vig, and Volpin (2015), Serfling (2016)), production flexibility (Reinartz and Schmid (2016)), investment flexibility (Barry et al. (2022)), and pricing flexibility (D'Acunto et al. (2018)).

The definition of flexibility implies that a flexible firm would respond well to economic shocks. Facing a negative shock, such as the COVID pandemic, a flexible firm can scale down its operation more easily to reduce loss or even adapt to win market shares from its competitors. Facing a positive shock, such as a technological advancement, a flexible firm can more easily adjust its production to the new technology to gain more.

2.2. How does D&I affect corporate flexibility?

From a labor management perspective, striving to be diverse and inclusive puts an extra constraint on a firm's hiring, firing, and promoting decisions. In the statistical discrimination theory of Phelps (1972) and Arrow (1972), group characteristics like race and gender are informative about individuals' true quality that is unobservable to the firm. Consequently, to be diverse and inclusive, the firm must collect additional costly information about individuals' true quality before each personnel decision, making such decisions less flexible. For promoting decisions specifically, Athey, Avery, and Zemsky (2000) theorize that optimal promotion within a firm often favors employees from the majority group because mentoring them for upper management positions is less costly. Thus, striving to be diverse and inclusive means that the firm often deviates from optimal promotion choices. As Fryer Jr and Loury (2013, p. 749) put it, diversity and inclusion goals "cannot be achieved without altering selection standards, distorting human capital investment decisions, or both." Together, these theories predict that a diverse and inclusive firm has less flexibility with personnel decisions, and hence less operating flexibility.

There are likely organizational costs associated with D&I as well. First, many firms implement D&I trainings, for example, to circumvent unconscious bias and sexual misconduct, which take time away from the firms' main business operations. Second, because building a culture of diversity and inclusion likely requires substantial investments (Gorton and Zentefis (2020)), a firm with a strong D&I culture may be reluctant to lay off employees at the risk of destroying its past investments in D&I and incurring future costs of rebuilding such a culture. Third, to be diverse and inclusive, a firm likely has metrics and processes to ensure progress on D&I, which could slow

things down inside the firm. For example, many companies adopt the Rooney rule, a policy to consider at least one minority candidate per job opening, which requires more time and effort with hiring practices. D&I-related processes might even directly restrict a firm from flexibly firing minority employees and thus reduce the firm's operating flexibility to scale down in bad times.

On the other hand, a D&I firm might be more flexible and adaptable to economic shocks because a diverse and inclusive workforce could gather and process information better. Women and minorities often have a more diverse network of information (Ibarra (1993), Ibarra (1995), Ibarra (1997)), allowing their employers to gather more information under changing economic conditions. Early lab evidence from Hoffman (1959) also indicates that diverse groups process information better than homogenous groups. The effect is large: Hoffman and Maier (1961) document that 65% of the diverse groups "produced high quality solutions… compared to only 21% of the homogeneous groups." Similarly, Nemeth (1986) documents that the mere presence of minorities increases critical thinking and creativity for group performance. These findings suggest that D&I could help a firm gain an advantage in gathering and processing information. Such an information advantage likely matters more when information becomes more valuable, such as during periods with large economic shocks.

In addition, successfully managing diversity requires both a firm's organizational structure and its employees to become more flexible. Cox and Blake (1991) argue that managing diverse groups forces a firm to broaden its policies and procedures, thus becoming more fluid and adaptable to changes. Cox and Blake also argue that the exposure to diverse groups in a firm could increase its employees' and management's tolerance to different viewpoints, and so, increase their openness to new ideas and changes.

Lastly, diversity and inclusion (D&I) could increase a firm's labor supply, allowing for more flexibility to expand when needed. Being diverse and inclusive means that a firm does not tolerate taste-based discrimination (Becker (1957)) in its hiring practices, thereby broadening the firm's search for talent. Furthermore, D&I could make a firm appear attractive to prospective employees. For example, in an experiment by Madera (2019), women are more likely to apply to a firm that they perceive to be more diverse and inclusive of women. Consequently, D&I could enhance a firm's access to labor and thus increase its flexibility to expand during economic booms. The larger labor supply, however, is unlikely to help a firm during economic busts.

Overall, it is an empirical question whether D&I helps a firm operate more flexibly and thus respond better to economic shocks. I state these hypotheses formally in null forms below:

H1*: "D&I firms and non-D&I firms do not differ on a measure of flexibility."*

H2*: "Following a large economic shock, a D&I firm does not outperform a non-D&I firm."*

3. Measuring D&I

In this section, we describe the data and the methodology we use to measure diversity and inclusion (D&I) at the firm-year level. We then present summary statistics and tests to validate the measure.

3.1. Data

We obtain employee reviews from Glassdoor.com, a career intelligence website. Glassdoor was launched in 2008, aiming to collect anonymous reviews from employees about employers. Glassdoor quickly became so popular that it started to provide job search services as well and became the number 2 job search site by user base in 2017. Glassdoor employs many mechanisms to control the quality of reviews. For example, Glassdoor's give-to-get policy requires each of its users to contribute to the website before accessing others' reviews on the website. This policy incentivizes more people to write reviews, especially those with non-extreme views, thus making Glassdoor reviews less polarized (Marinescu et al. (2018)). The website also claims to review every contribution by its users, to ensure that its reviews are helpful, authentic, and balanced. In this paper, I obtain 10.4 million Glassdoor reviews for over 300,000 employers as of May 2021.

A typical Glassdoor review contains a review title, date written, employee title, employee status (former vs. current), city and state of location, years in the company, numerical ratings for overall evaluation, work-life balance, culture, career, compensation, management, and text fields containing the pros and the cons of working at the company. In September 2020, Glassdoor launched a new rating called Diversity and Inclusion on its website, the first of its kind.[3] This feature allows each Glassdoor user to anonymously rate his or her employer's D&I practices on a scale of one to five stars. As of May 2021, over 1.1 million reviews have a non-missing D&I rating. Appendix A shows an example of a typical review.

[3] See https://www.glassdoor.com/blog/diversity-inclusion-products/.

While the majority of reviews have all the typical information listed above, only the overall rating, the pros and cons sections, and the review date and title are mandatory for every review. The other ratings are not mandatory: only 79.7% of the reviews have all of them available (see the Internet Appendix Table IA1 for the review-level summary statistics). Thus, I rely on mostly the overall rating, review title, and the pros and cons texts as the main input into my model to predict the D&I rating for all reviews, and I use the other information in a review whenever possible.

3.2. Measuring D&I with BERT

While the newly introduced D&I rating by Glassdoor is only available since 2020, it is possible to construct a D&I rating for all reviews since 2008. The idea is to extrapolate the D&I rating back in time using a model relating the D&I rating to each review's written texts and other ratings, which have been available since 2008.

To allow my model to understand review texts well, I use BERT (Bidirectional Encoder Representations from Transformers), a breakthrough model in natural language processing. Google introduced BERT in 2018 to understand search queries better, and the model even outperformed humans in reading comprehension (Devlin et al. (2018)). BERT represents each word in a sentence and the whole sentence by different vectors and allows these vectors to interact and change. As the model updates itself in predicting a word based on its surroundings and a sentence based on its preceding sentence, BERT can understand each word in its context and the sentence as a whole. More importantly, Google pre-trained BERT using a large corpus (3.3-billion-word corpus from English books and Wikipedia), so BERT has gained significant knowledge of how English words and sentences are often represented together. Consequently, by fine-tuning the BERT model, i.e., allowing the model's weights to change, to fit better to my setting of predicting the D&I rating based on review texts, I can transfer the model's general knowledge of the English language to the specific context of employee reviews.

To apply the BERT model to predict the D&I rating using each review's texts and other ratings, I proceed in three steps. First, I combine the relevant information in each review into a single text that BERT can process. In particular, for the example review in Appendix A, the single text to feed into the BERT model is "Pros are as follows: *It is LGBTQIA+ friendly, they don't discriminate based on appearances…* Cons are as follows: *Upward mobility is very difficult within the same*

store and varies based on position… I give an overall rating of *3.0*, a work/life balance rating of *3.0*…". Appendix A shows this combined text in full.

Second, I feed the combined text into a simplified version of BERT, called distilled BERT (Sanh et al. (2019)), to reduce the computational burdens. In doing so, I instruct the model to predict the D&I rating in a classification task with five classes representing five possible ratings: one to five stars. I train the model on 1.19 million reviews with a non-missing D&I rating as of May 2020, using a standard split of 25% for the testing sample and 75% for the training sample.

Third, I use the trained model to predict the D&I rating for all the 10.4 million reviews available since 2008. Because not all reviews contain all the other ratings, I train two separate models for prediction: one using the pros, cons, title, and overall rating as input, which are available for all reviews (simpler model), and another using all the available information, including other ratings (fuller model), whenever possible. The out-of-sample D&I rating for each review is then the D&I rating predicted by the fuller model whenever the data allow, and the simpler model's prediction otherwise. The Internet Appendix Table IA1 Panel C shows that the classification accuracy of the simpler model is above 55%, and above 63% for the fuller model. The correlation between the predicted D&I rating and the actual D&I rating is as high as 0.78.

Given the predicted D&I rating for each review, I calculate each firm's D&I rating in a year by averaging the predicted D&I rating across reviews in that firm-year.

3.3. Validating the measure

I validate the D&I rating at both the review level and the firm-year level.

3.3.1. At the review level

My model predicts a D&I rating for every review in over nine million employee reviews by training on over one million reviews in which employees provide an actual D&I rating. However, if employees misunderstand or confuse D&I with other factors like job satisfaction, the model may not accurately capture D&I.

To address this concern, I adopt the approach of Acikalin et al. (2022) to identify the most informative words about the model-based D&I rating across reviews. Table IA2 in the Internet Appendix presents the top 30 words in the pros and cons sections of reviews that are most highly associated with the D&I rating. The pros section includes words like *inclusive*, *listen*, *diverse*, and

diversity, while the cons section includes words like *racist*, *discrimination*, *sexual*, *bullying*, and *harassment*. These words indicate that the D&I rating does indeed capture the concept of D&I.

In addition, I assess how well my model captures D&I relative to human ratings. While the model learns from human ratings, its training on multiple reviews may allow it to filter out idiosyncratic noise, allowing it to zoom in discussions of D&I issues more than human ratings. To evaluate this possibility, I adopt Briscoe-Tran (2022)'s methodology, constructing a D&I measure for each review based on the frequency of D&I keywords in the pros relative to the cons section. Because this word-count approach targets D&I words directly, it is a reasonable benchmark for measuring D&I. I find that the word-count measure's correlation with my model-based D&I rating is over 0.34, while its correlation with the human ratings is 0.29. This result implies that my model-based D&I rating more closely aligns with written reviews of D&I issues than the actual employee-provided D&I rating.

3.3.2. At the firm-year level

I take the average estimated D&I rating across reviews in each firm-year to be my main D&I measure at the firm-year level.[4]

Ideally, to validate this measure, I need indicators of what D&I practices are like inside a firm. But such indicators are rare. External ratings of a firm's D&I practices are available for a limited subset of firms, such as the Refinitiv D&I rating, but this external rating captures mostly corporate policies on D&I and what firms describe their D&I practices to be, which are prone to window-dressing incentives (e.g., Delmas and Burbano (2011), Marquis, Toffel, and Zhou (2016), Fabrizio and Kim (2019), Briscoe-Tran (2022), Baker et al. (2022)).

Luckily, one good validation indicator for a firm's internal D&I practices is available. That is the list of 100 Best Workplaces for Diversity, which the Fortune magazine and the Great Place to Work Institute collaborated to construct from an anonymous survey of minorities within a firm, such as women, people of color, LGBTQIA+ people, older employees, and employees with disabilities. While the list was discontinued in 2020, its values for the years between 2015 and 2019 could help validate my D&I measure, akin to an out-of-sample test.

[4] I also consider the dispersion in the D&I rating across reviews in a firm-year as a potential indicator for poor D&I practices. While this measure is negatively related to a firm's future D&I-related outcomes, I find in un-tabulated tests that this relationship is not robust after controlling for my main D&I measure.

Table 2 Panel A shows that my D&I measure strongly predicts the chance of a firm landing in Fortune's Best Diversity list one year ahead (and two and three years ahead as well in un-tabulated analyses). The logit coefficient on the D&I measure is 2.05, statistically significant at the 1% level in column (1), indicating that the odds of a firm's landing in Fortune's Best Diversity list increases by 7.7 times as a firm's D&I rating increases by one star. It remains large and statistically significant after controlling for industry and year fixed effects (columns (2) to (8)), firm characteristics (column (4) and (5)), and the lagged indicator of the Best Diversity list (column (3) and (5)). The Refinitiv rating of corporate D&I policies also predicts the likelihood of a firm landing in the Best Diversity list, but the predictive power of my D&I measure remains statistically significant after controlling for the Refinitiv rating (column (6)). The economic magnitude of the coefficient on my D&I rating is four times larger than that of the Refinitiv rating.[5] Finally, despite the D&I rating's high correlation with the overall rating and the culture rating on Glassdoor, my D&I measure remains a strong and significant predictor of the Best Diversity list after controlling for these other ratings (columns (7) and (8)).

One concern with the above predictive test is a potential look-ahead bias, given that my D&I measure is estimated using a model trained on 2020 data and the predictive test is performed on data predating 2020. However, this concern is limited because the underlying input for the prediction are reviews written before the measurement of the predicted outcomes. For example, suppose that all the reviews in 2018 mention nothing about D&I practices, then the model, regardless of how it is trained, will predict the same D&I rating across firms in 2018, leaving it no predictive power for any D&I-related outcomes in 2019. Nonetheless, to be sure, I examine whether the D&I rating's predictive power remains unchanged in predicting D&I-related outcomes after 2021, for which there is undoubtedly no look-ahead bias for a model trained on 2020 data. While the Fortune's Best Diversity List was discontinued in 2020, Forbes Magazine recently started constructing a similar list with a similar survey methodology. Thus, I collect the top 100

[5] The odds would increase by exp((26.5-8.5)*0.051)=2.50 times as the Refinitiv rating moves from its 25th percentile to its 75th percentile. The same number for my D&I rating would be exp((3.855-3.036)*2.852) = 10.33.

firms on this Forbes' Diversity list for 2023 and find in Table 2 Panel B that the D&I rating measured in 2019 strongly predicts the chance of a firm landing on this list as well.[6] [7]

Next, as a comparison, I investigate whether my D&I measure predicts a firm's likelihood of landing in *another* Diversity and Inclusion list that is *not* based on how employees view D&I practices. This other list is the DiversityInc Top 50 ranking, which is based on a survey that firms voluntarily participate to describe their D&I performance.[8] Since this alternative Best Diversity list relies on firms' own disclosure, it is more likely to be correlated with the Refinitiv D&I rating than my rating. I find that to be the case, as shown in Table 2 Panel C. My D&I measure does not significantly predict the likelihood of a firm's landing in the DiversityInc Top 50 list after controlling for the Refinitiv D&I rating (column (6)). However, my D&I measure still significantly predicts the likelihood of a firm's landing in the DiversityInc Top50 in other specifications. Together, these results imply that while the Refinitiv D&I rating better captures a firm's own disclosure of D&I practices, my D&I measure better captures the firm's grassroot D&I practices.

Finally, I evaluate the persistence in my D&I measure as a way to validate it. As Gorton and Zentefis (2020) theorize, a firm's internal practices, i.e., corporate culture, are likely very hard to change over time, so a measure of internal D&I practices should be persistent. I find that to be the case. Table 2 Panel D shows that the autocorrelation in the D&I measure is 0.440 (column (1)). The autocorrelation coefficient remains large and statistically significant after controlling for industry and year fixed effects (column (2)). It increases to 0.714 when I restrict my sample to firm-years with above-median number of reviews, where idiosyncratic noise is likely less (column (4)). Most importantly, even after controlling for other ratings on Glassdoor, the lagged D&I rating still significantly predicts itself going forward (columns (3) and (6)). These results imply that not only my D&I measure shows strong persistence within each firm, but its persistence also captures unique information beyond other ratings on Glassdoor.

Overall, my D&I measure appears to capture firms' D&I practices well.

---

[6] Forbes publicly shows the Best Diversity list for the most recent year (2023 at the time of this writing) at: https://www.forbes.com/lists/best-employers-diversity/?sh=76762f306468.

[7] The D&I rating also aligns well with D&I outcomes at the state-year level that are monitored by the US Equal Employment Opportunity Commission (EEOC). The correlation between the D&I rating, aggregated to the state-year level, and the frequency of EEOC discrimination charges per capita for US states between 2008 and 2020 is -0.189.

[8] https://www.diversityinc.com/diversityinc-top-50-methodology/

3.4. Summary statistics

While I have reviews covering over 300,000 employers listed on Glassdoor, in this paper I focus on a sample of the largest publicly listed companies in the US because these firms report abundant data on operating performance. In particular, I restrict my sample to all US-domiciled firms that were ever listed during 2008-2020 with at least 100 reviews available on Glassdoor as of July 2020. The final sample consists of 2,113 firms, with 2,617,482 reviews between Glassdoor inception in 2008 and May 30, 2021. The Internet Appendix IA1 describes in detail how I construct my sample and match that to other databases like Compustat for accounting data.

I summarize the data at the firm-year level in Table 1. Panel A shows that the average D&I rating ranges between 1.33 and 5.00 when moving from the 1st percentile to the 99th percentile, with an average value of 3.44. Other ratings show a similar range. Panel B shows that the D&I rating has a significant correlation with other firms' characteristics, such as 0.094 for size, 0.072 for sales growth, 0.185 for Tobin's q, -0.090 for debt ratio, and 0.137 for cash-to-assets ratio. These correlations mean that bigger firms, growth firms, firms with higher valuation, firms with lower leverage, and cash-rich firms tend to have better D&I practices from employees' perspective. The D&I rating also has a positive correlation with board gender diversity (0.186), top management gender diversity (0.139), and the overall D&I score by Refinitiv (0.254), and a negative correlation with the Refinitiv D&I controversy score (-0.077), although these diversity data are only available for a small subset of firms (under one third of my sample).

In light of these correlations, I regress the D&I rating on various sets of fixed effects and firm characteristics. The Internet Appendix Table IA4 confirms that the above correlations remain quantitatively similar. In addition, it shows that industry and year fixed effects explain up to 9.9% of the variation in the firm-year level D&I rating, while firm fixed effects explain up to 38%.

Finally, the Internet Appendix Figure 1A shows that all the ratings on Glassdoor at the firm-year level, including the D&I rating available since 2020 and the predicted D&I rating since 2008, have a bell-shaped distribution, suggesting that Glassdoor reviews are indeed balanced with few extreme reviews.

Taking the measure as given, Figure IA2 in the Internet Appendix indicates that D&I practices in the average firm in my sample have improved over time, with the measure increasing from 3.3 in 2013 to 3.7 in 2020, coinciding with the trend in the Refinitiv rating of corporate D&I policies. The D&I measure also indicate that while Southwest Airlines and Facebook Inc. have the best

D&I practices during 2016-2020, Union Pacific, a railroad operating company, and PacifiCorp, an electric power company, have the worst D&I practices (Internet Appendix Table IA3).

4. Do D&I firms exhibit lower flexibility?

In this section, we test whether D&I firms exhibit more or less flexibility than other firms, using a theoretically motivated measure of flexibility from the literature. First, we provide cross-sectional evidence on the relationship between a firm's D&I rating and its flexibility. Second, we provide causal evidence on how a firm's flexibility changes following a plausibly exogenous shock to D&I.

4.1. Cross-sectional evidence.

The literature has struggled to measure a firm's flexibility (Chen, Kacperczyk, and Ortiz-Molina (2011)). Fortunately, Gu, Hackbarth, and Johnson (2018) derive a theoretically motivated measure of flexibility based on a firm's historical range of operating cost margins relative to its fluctuations in productivity (measured by sales over assets). The idea is that as productivity shocks hit firms, a less flexible firm would wait longer to adjust its operations, allowing the shocks to bring its operating cost margins to more extreme levels, i.e., a wider range.

For example, when a positive shock occurs to a firm, its profit margins will increase. A flexible firm would then adjust to increase production to take advantage of the shock. As production increases, declining returns to scale would result in lower profit margins, leaving the firm's profit margins (as well as cost margins) more stable. By contrast, an inflexible firm would fail to increase production, leaving its profit margins at a high level. The argument reverses when a negative shock occurs. Thus, as shocks affect firms' profitability, an inflexible firm will end up with a wider range of profit and cost margins.

Consequently, I follow Gu, Hackbarth, and Johnson (2018) to measure corporate flexibility. Since computing this measure requires accounting data over many years, it prevents powerful tests in the time-series. So, I start the analysis with the following cross-sectional regression model, basically testing whether firms with a better D&I rating tend to have lower or higher flexibility:

$$FLEX_i = \beta_0 + \beta_1 * D\&I_i + \mathrm{B}_2 * X_i + v_j + \epsilon_i$$

Where $FLEX_i$ is the inverse of INFLEX, defined as the range of operating cost margins over the standard deviation of log changes in sales over assets, using accounting data for firm i from 2009 up to 2022, the time of Compustat data download. $D\&I_i$ is the average of firm i's D&I rating

based on Glassdoor reviews during the same time frame, though only up to 2021, when I downloaded Glassdoor data. While Glassdoor reviews started in 2008, I drop this inception year because reviews were only available for half a year then. X_i denote various control variables, like size, calculated as the average value across all the years from 2009 up to 2022 for each firm. v_j refers to the Fama-French 48 industry indicators, and ϵ_i is the error term. To reduce noise, I winsorize all ratio variables at the 1st and 99th percentiles, and drop firms with fewer than five years of available cost margins data. [9] [10]

The results in Table 3 show that D&I firms exhibit significantly lower flexibility than other firms. The coefficient on the D&I rating in column (1) indicates that a firm with one star higher in the D&I rating operates on a range of operating cost margins that is over 36% standard deviation larger (0.691/1.907). This coefficient is statistically significant at the 1% level, and remains significant at the 1% level even after controlling for industry fixed effects (column (2)), and many firm characteristics such as financial flexibility (size, leverage, and cash), labor stock (number of employees relative to total assets), fixed assets, and Tobin's Q (column (3)). Since corporate governance likely matters for a firm's flexibility, I control for proxies of governance, including institutional ownership percentage and concentration, and inside ownership as well, but the coefficient on the D&I measure remains large and significant (column (4)). The result is similar when I control for Bebchuk, Cohen, and Ferrell (2009)'s measure of governance (entrenchment index) too, but I do not include it in the main tests because doing so significantly lowers the sample size. Finally, the coefficient on the D&I measure remains strong and significant when I control for a host of diversity measures at the board and top management level in column (5), such as gender mix, nationality mix, tenure diversity, age diversity, and the network size of a firm's board and top management, all from the BoardEx database. These results suggest that the effect of workforce D&I on a firm's flexibility is separate from that of board and top management diversity.[11]

[9] While measuring FLEX over a shorter window, say five years, may be inadequate in capturing flexibility, doing so could allow for analyzing a lead-lag relationship between D&I and flexibility. In un-tabulated tests, I find that D&I in year t is negatively and significantly associated with a firm's flexibility, measured over the t to t+5 period.

[10] I use Compustat annual files, because Das, Shroff, and Zhang (2009) find that firms use accounting discretion to adjust 4th quarter earnings upwards when the other quarters have low earnings, creating excess volatility in quarterly profit and cost margins. In un-tabulated tests, my results remain similar when FLEX is measured with quarterly data.

[11] The results hold similarly after I control for a measure of corporate D&I policies from Refinitiv. I do not include it in the main table because it is only available since 2016 and for a significantly smaller number of firms.

4.2. Within-firm evidence from a D&I shock.

The negative association between a firm's D&I rating and its flexibility need not imply causation. To investigate whether there is a causal relationship between D&I and flexibility, I study a plausibly exogenous shock to firms' D&I practices. The shock happened in 2013 when there was an unexpected court ruling related to D&I.

In the United States, employers could be held accountable for workplace harassment if the harasser holds a position of authority over the victim, as mandated by the Title VII of the Civil Rights Act of 1964. However, in July 2013, the 7th Circuit Court, which establishes legal precedents for cases in Illinois, Indiana, and Wisconsin, issued an unexpected ruling in the case of *Lambert v. Peri Formworks Sys., Inc*. This ruling held the employer responsible for sexual and racial harassment even when the harasser was merely a co-worker of the victim. As a result, firms located in these three states faced an increased risk of harassment lawsuits compared to firms located elsewhere after the ruling. Notably, the court held the employer liable despite the existence of the employer's policies for handling harassment complaints. This meant that treated firms could not simply rely on adding more policies to mitigate the increased legal risk. Instead, they had a stronger incentive to genuinely enhance their diversity and inclusion (D&I) practices.

If firms improve their D&I practices after the ruling, a causal relationship between D&I and corporate flexibility would imply a decline in these firms' flexibility. In addition, because the ruling was more likely to affect firms with existing poor D&I practices, such as firms with racial and sexual discrimination issues, I expect the improvement in D&I and the decline in corporate flexibility, if present, to be concentrated among those firms.

In a difference-in-differences design, I compare treated firms (headquartered in the three states influenced by the 7th Circuit Court) with other firms located in the US. For each firm, I calculate the flexibility measure over the two five-year periods before and after the court ruling: 2008-2012 and 2013-2017. I then observe how firms' flexibility changed from one period to the next and how the firms' D&I rating changed as well. I do so for two different subsamples based on prior D&I rating: high (above median) and low (below median). The D&I rating in 2012, right before the court ruling, serves as the prior rating, against which I calculate subsequent year changes.

Table 4 depicts the results. In the overall sample, the treated firms experienced an improvement in their D&I ratings and a decline in their flexibility after the court ruling, relative to control firms (columns (1) and (2)). The results concentrate among the subsample with a low prior D&I rating,

where the increase in D&I and the decrease in flexibility were larger and highly significant (columns (3) and (4)). By contrast, the high prior D&I rating subsample shows insignificant changes in the D&I rating and the flexibility measure (columns (5) and (6)). These contrasting patterns highlight that flexibility declines only when D&I improves, consistent with a causal relationship between D&I and flexibility. Regarding economic magnitudes, since the standard deviation of FLEX when measured over the 5-year window is 5.183, the results suggest that after the court ruling, flexibility declined by up to 0.31 standard deviation (1.611/5.183).

Graphical evidence in Figure 1 is consistent with the above analyses: the D&I rating increased while flexibility declined after the court ruling for the affected firms relative to other firms, and these effects concentrated among firms with a low D&I rating before the court ruling. In addition, the graphs suggest that there is no pre-trend in the outcome variable, flexibility, before the court ruling. For the D&I rating, there is no statistically significant pre-trend in the full sample, but there appears to be some pre-trend for the subsample of firms with a low prior D&I rating. However, the measurement for the D&I rating before the shock is less reliable due to the limited data coverage of Glassdoor in its early years. Nonetheless, to ensure that the result is not driven by the pre-trend, I perform a nearest neighbor matching procedure to find up to three control firms for each treated firm based on their D&I ratings in the two years before the court ruling, as well as size and industry. In this matched sample, the pre-trend disappears while the main results remain similar, as shown in the Internet Appendix Figure IA3.

5. Mechanisms

Diversity and inclusion (D&I) can impede corporate flexibility through two potential channels. The first channel relates to the *workforce*, where D&I considerations introduce complexities in hiring and firing processes, limiting a firm's ability to swiftly manage its employees. The second channel involves *efficiency*, wherein D&I firms encounter communication challenges among diverse groups, resulting in slower decision-making and less efficient adaptation to changes.

To investigate these channels, we examine how D&I firms handle their workforce and operating efficiency following an unexpected economic shock. In my sample period, the COVID-19 crisis in 2020 emerged as an ideal and substantial unanticipated shock, providing an opportunity to observe how firms responded to large changes in economic conditions.

5.1. Do D&I firms exhibit lower flexibility during the shock?

First, we test whether D&I firms exhibit lower flexibility relative to other firms during the COVID crisis. If D&I firms have less flexibility than others, they should experience a larger decline in operating performance than other firms. I measure operating performance by return on assets (ROA). I do not study stock returns because flexibility is about *short-term* adaptation to changes while stock returns often reflect *long-term* expectations of a firm's performance.

We regress the change in a firm's ROA between 2019 and 2020 on the firm's D&I rating and other firm characteristics in 2019 and interact these variables with an indicator of high exposure to the COVID shock (based on how much a firm's industry relies on in-person interactions, following Koren and Peto (2020)). Table 5 shows the results. To save space, I only show the coefficients on the exposure measure and its interactions with firm characteristics.

Column (1) confirms the literature's finding that highly exposed firms experienced a larger decline in profits (2.6% of total assets) during 2020, relative to other firms. The decline was more pronounced for firms with a high D&I rating. The coefficients in column (2) indicate that among firms with a high COVID exposure, ROA declined by 3.2% more in firms with one star higher in the D&I rating. This estimate is statistically significant at the 1% level, and remains so after controlling for the interactions of the high exposure indicator with proxies for financial flexibility (column (3)), proxies for corporate governance and assets attributes (column (4)), labor intensity (employee count over total assets), labor skill (reliance on H1-B workers) and advertising intensity (column (5)), and industry fixed effects (column (6)). Advertising intensity and industry fixed effects likely capture how important a large set of customers is to a firm, so the results so far are unlikely because D&I firms are often firms with a large customer base and thus most susceptible to COVID-induced social distancing.

Similarly, the results hold after controlling for a firm's reliance on H1-B workers (high-skilled immigrants), so they are unlikely because D&I firms may rely more on these workers, who experienced a tightening of immigration policies during the COVID period. The results also remain the same after controlling for diversity indicators at the board and top management level and their interactions with the exposure measure (un-tabulated to save space).

Overall, I find that D&I firms did not adapt as well as other firms to the COVID shock to protect their performance, implying lower flexibility for D&I firms. The effect of D&I on

flexibility in responding to the COVID shock appears to be distinct from the effect of financial slack, labor intensity, corporate governance, board diversity, and other firm characteristics.

5.2. Which channels: workforce or efficiency?

Next, I investigate whether D&I firms suffer more during the COVID period due to their less flexible workforce management or their less flexible operating efficiency management. To do so, I conduct similar tests to the previous section but instead focus on workforce cuts and measures of efficiency. For efficiency, I study both cost efficiency and revenue efficiency. The cost efficiency measures include cost ratios like selling, general, and administrative expenses relative to sales. The revenue efficiency measures include sales per employee, sales growth, and assets turnover.

Table 6 shows the results. Panel A column (1) indicates that firms with a high exposure to the COVID shock had a significant 6.5% larger cut in their total number of employees in 2020. When I allow the exposure measure to interact with firms' D&I ratings and other characteristics before 2020, as shown in columns (2) to (6), there is no significant coefficient on the interaction between the D&I rating and the exposure measure, suggesting that the COVID-induced workforce cut was not more pronounced for D&I firms relative to other firms. Therefore, the overall decline in D&I firms' performance during the COVID crisis was unlikely due to these firms failing to flexibly cut their workforce in response to the shock.

In Panel B, I focus on changes in a firm's cost ratios following the COVID shock, particularly on selling, administrative, and general (SG&A) expenses relative to sales. Column (1) indicates that firms with a high COVID exposure incurred a significantly higher increase in SG&A expenses. Columns (2) to (6), however, indicates that such an increase was not more pronounced for D&I firms relative to other firms. Un-tabulated analyses find similar results for capital expenditure as well. Thus, differential changes in expense ratios do not explain why D&I firms have lower operating performance than other firms during the COVID shock.

Panel C, by contrast, shows a different pattern for efficiency measures regarding revenue. Column (1) confirms the literature's finding that highly exposed firms experienced a larger decline in sales growth (5.6%) during 2020, relative to other firms. The decline was more pronounced for firms with a high D&I rating, by a margin of 7.5% more in firms with one star higher in the D&I rating. This estimate is statistically significant at the 1% level, and remains so after controlling for all the control variables as before.

Panel D paints a similar picture. Sales per employee declined more for D&I firms relative to other firms with a high COVID exposure. Column (2) estimates indicate that comparing among firms with a high COVID exposure, a 4-star D&I rated firm experienced a 26,000 dollars larger decline in revenue per employee than a 3-star rated firm. This estimate is equivalent to over 35% of a standard deviation change in sales per employee in my sample. The estimate remains statistically significant at the 5% level after controlling for different trends among firms with different size, cash, debt, labor intensity, labor skill (H1-B reliance), and other firm characteristics (columns (3) to (5)). Even after controlling for differential industry trends (column (6)), the estimate remains economically sizable at over $20,000 per employee and statistically significant at the 10% level. Un-tabulated analyses show similar results for another measure of revenue efficiency: sales over assets, or assets turnover. The results also remain the same after controlling for diversity indicators at the board and top management level (un-tabulated to save space).

In conclusion, the results suggest that D&I firms exhibit lower flexibility in protecting their performance during the COVID shock, and the main reason for this lower flexibility appears to be a larger decline in revenue efficiency rather than a failure to manage costs or workforce cuts.

6. Robustness and discussion

In this section, we discuss potential caveats and the robustness of the main results in the paper.

Flexibility vs. operating leverage

While we find that D&I firms exhibit lower flexibility according to the flexibility measure in Gu et al. (2018), an alternative interpretation is that D&I firms simply have more operating leverage, i.e., higher fixed costs, leading to a lower ability to adjust costs in responding to changes. However, Gu et al. argue that their range-based measure captures flexibility in a way that is distinct from operating leverage. The idea is that even a firm with high fixed costs (high operating leverage) can still have a small range of cost margins if the firm is flexible enough to counter fluctuations in its cost margins. Empirically, Gu et al. find their inflexibility measure differs from operating leverage proxies and is correlated with adjustment cost proxies like a firm's investment-Q sensitivity. In addition, my results in Section 5 show that D&I firms exhibit lower flexibility not because of their failure to cut costs, but because of their larger declines in revenue efficiency during a crisis.

Robustness to other determinants of flexibility

Existing literature shows that there are many potential determinants of corporate flexibility: financial slack, fixed assets, and labor-related factors (Barry et al. (2022)) such as labor union (Chen, Kacperczyk, and Ortiz-Molina (2011)). So far, all the results in the previous sections are robust to controlling for cash and debt ratios as proxies for financial slack; property plant and equipment to total assets as a measure for fixed assets; and labor intensity and high-skilled labor ratio as a measure for labor-related factors. While I cannot control for labor union directly due to the lack of firm-level union data, I find that the results are robust to controlling for industry trends and industry fixed effects, which are known to capture variations in labor union coverage (Chen et al. 2011). In addition, after a shock to specifically D&I, which allows me to rule out unrelated factors like labor union, I show that a firm's D&I rating increases while flexibility declines, consistent with an effect of D&I on flexibility.

Caution in interpreting the court ruling results.

Section 4.2 shows that a firm's flexibility declined after its D&I rating increased due to an exogenous shock to D&I (the 2013 court ruling on racial and sexual harassment), thus ruling out omitted variable concerns. However, one cannot attribute all the decline in flexibility to the increase in the D&I rating. Because, while all the changes a firm made to improve its D&I practices may restrict its flexibility, only a fraction of the changes may succeed in improving its D&I rating. Nonetheless, as long as the changes firms made were motivated by the D&I-related court ruling, we can still conclude that D&I considerations do reduce flexibility.

Caution in interpreting the COVID results.

Section 5 utilizes the COVID crisis as a significant and unanticipated shock to study the mechanisms for how D&I practices may lower a firm's flexibility. The conclusion is that D&I firms exhibit lower flexibility mainly because they face more frictions with operating efficiency, rather than workforce management, in responding to the shock. One caution is that this conclusion may not generalize to other settings, i.e., other types of shocks. Nevertheless, the findings so far still allow for the interpretation that the efficiency channel does play a role.

7. Conclusion

In this paper, we examine the impact of diversity and inclusion (D&I) practices on corporate flexibility. Using a unique employee rating system introduced in 2020, we analyze data from thousands of companies dating back to 2008 and find that diverse and inclusive firms (D&I firms)

exhibit lower flexibility. Exploration of mechanisms suggests that D&I firms have lower flexibility due to their worsened operating efficiency in response to unexpected economic shocks.

The findings imply that there exists a downside to having good D&I practices for a corporation: lower flexibility. Thus, the study informs business practitioners about the cost of diversity and inclusion, allowing them to make more informed decisions about D&I practices within their organizations. Nonetheless, the findings do not mean that having good D&I practices is not optimal for firms, because there could be many potential benefits of D&I that future studies can examine, such as higher innovation, more efficiency during normal times, and better risk management.

While studying the effect of D&I on other aspects of a firm is beyond the scope of this paper, the paper provides a novel firm-level measure of D&I practices that future research can use to study those other aspects. For example, because the literature documents that firms with lower operating flexibility have a stronger precautionary motive to hold more cash and use less debt, future research can explore the impact of D&I practices on a firm's cash and debt policies. Operating flexibility also has implications for a firm's risk and cost of capital, so future research can examine how D&I affects risk and return. Finally, a firm-level D&I measure can also help study the effect of policy interventions like board gender quotas, public outrages like the George Floyd killing, or top management on workforce diversity and inclusion.

References

Acikalin, Utku U., Tolga Caskurlu, Gerard Hoberg, and Gordon M. Phillips, 2022, Intellectual Property Protection Lost and Competition: An Examination Using Machine Learning, National Bureau of Economic Research.

Alesina, Alberto, Reza Baqir, and William Easterly, 1999, Public Goods and Ethnic Divisions, *The Quarterly Journal of Economics* 114, 1243–1284.

Alesina, Alberto, and Eliana La Ferrara, 2002, Who Trusts Others?, *Journal of Public Economics* 85, 207–234.

Arrow, Kenneth J., 1972, Models of Job Discrimination, *Racial Discrimination in Economic Life*.

Athey, Susan, Christopher Avery, and Peter Zemsky, 2000, Mentoring and diversity, *American Economic Review* 90, 765–786.

Avery, Derek R., and Patrick F. McKay, 2006, Target Practice: An Organizational Impression Management Approach to Attracting Minority and Female Job Applicants, *Personnel Psychology* 59, 157–187.

Baker, Andrew, David F. Larcker, Charles McClure, Durgesh Saraph, and Edward M. Watts, 2022, Diversity Washing, *Chicago Booth Research Paper*.

Barry, John W., Murillo Campello, John R. Graham, and Yueran Ma, 2022, Corporate Flexibility in a Time of Crisis, *Journal of Financial Economics* 144, 780–806.

Bebchuk, Lucian A., Alma Cohen, and Allen Ferrell, 2009, What Matters in Corporate Governance?, *The Review of Financial Studies* 22, 783–827.

Becker, Gary S., 1957, *The Economics of Discrimination* (University of Chicago Press).

Benson, Alan, Simon Board, and Moritz Meyer-ter-Vehn, 2023, Discrimination in Hiring: Evidence from Retail Sales, *The Review of Economic Studies*, rdad087.

Bernile, Gennaro, Vineet Bhagwat, and Scott Yonker, 2018, Board Diversity, Firm Risk, and Corporate Policies, *Journal of Financial Economics* 127, 588–612.

Bertrand, Marianne, and Esther Duflo, 2017, Field Experiments on Discrimination, *Handbook of Economic Field Experiments* 1, 309–393.

Briscoe-Tran, Hoa, 2022, Do Employees Have Useful Information About Firms' ESG Practices?, *Fisher College of Business Working Paper*, 21.

Butler, Alexander W., Erik J. Mayer, and James P. Weston, 2022, Racial Disparities in the Auto Loan Market, *The Review of Financial Studies*.

Chava, Sudheer, Wendi Du, Agam Shah, and Linghang Zeng, 2022, Measuring Firm-Level Inflation Exposure: A Deep Learning Approach, *Available at SSRN 4228332*.

Chen, Huafeng Jason, Marcin Kacperczyk, and Hernán Ortiz-Molina, 2011, Labor Unions, Operating Flexibility, and the Cost of Equity, *Journal of Financial and Quantitative Analysis* 46, 25–58.

Choi, Jung Ho, Joseph Pacelli, Kristina M. Rennekamp, and Sorabh Tomar, 2023, Do Jobseekers Value Diversity Information? Evidence from a Field Experiment and Human Capital Disclosures, *Journal of Accounting Research* 61, 695–735.

Cox, Taylor H., and Stacy Blake, 1991, Managing Cultural Diversity: Implications for Organizational Competitiveness, *Academy of Management Perspectives* 5, 45–56.

D'Acunto, Francesco, Ryan Liu, Carolin Pflueger, and Michael Weber, 2018, Flexible Prices and Leverage, *Journal of Financial Economics* 129, 46–68.

Das, Somnath, Pervin K. Shroff, and Haiwen Zhang, 2009, Quarterly Earnings Patterns and Earnings Management, *Contemporary Accounting Research* 26, 797–831.

Delmas, Magali A., and Vanessa Cuerel Burbano, 2011, The Drivers of Greenwashing, *California management review* 54, 64–87.

Denis, David J., 2011, Financial Flexibility and Corporate Liquidity, *Journal of Corporate Finance* 17, 667–674.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, 2018, BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding, *arXiv preprint arXiv:1810.04805*.

Duchin, Ran, Mikhail Simutin, and Denis Sosyura, 2021, The Origins and Real Effects of the Gender Gap: Evidence from CEOs' Formative Years, *The Review of Financial Studies* 34, 700–762.

Edmans, Alex, 2023, The End of ESG, *Financial Management* 52, 3–17.

Edmans, Alex, Caroline Flammer, and Simon Glossner, 2023, Diversity, Equity, and Inclusion, National Bureau of Economic Research.

Fabrizio, Kira R., and Eun-Hee Kim, 2019, Reluctant Disclosure and Transparency: Evidence from Environmental Disclosures, *Organization Science* 30, 1207–1231.

Field, Laura Casares, Matthew E. Souther, and Adam S. Yore, 2020, At the Table but Can Not Break Through the Glass Ceiling: Board Leadership Positions Elude Diverse Directors, *Journal of Financial Economics* 137, 787–814.

Frame, W. Scott, Ruidi Huang, Erik J. Mayer, and Adi Sunderam, 2021, The Impact of Minority Representation at Mortgage Lenders, *SMU Cox School of Business Research Paper*.

Fryer Jr, Roland G., and Glenn C. Loury, 2013, Valuing Diversity, *Journal of Political Economy* 121, 747–774.

Ghaly, Mohamed, Viet Anh Dang, and Konstantinos Stathopoulos, 2017, Cash Holdings and Labor Heterogeneity: The Role of Skilled Labor, *The Review of Financial Studies* 30, 3636–3668.

Gillan, Stuart L., Andrew Koch, and Laura T. Starks, 2021, Firms and Social Responsibility: A Review of ESG and CSR Research in Corporate Finance, *Journal of Corporate Finance* 66, 101889.

Gorton, Gary B, and Alexander K Zentefis, 2020, Corporate Culture as a Theory of the Firm, .

Green, T. Clifton, Ruoyan Huang, Quan Wen, and Dexin Zhou, 2019, Crowdsourced Employer Reviews and Stock Returns, *Journal of Financial Economics* 134, 236–251.

Griffin, Dale, Kai Li, and Ting Xu, 2021, Board Gender Diversity and Corporate Innovation: International Evidence, *Journal of Financial and Quantitative Analysis* 56, 123–154.

Gu, Lifeng, and Dirk Hackbarth, 2021, Inflexibility and Leverage, *Working Paper*.

Gu, Lifeng, Dirk Hackbarth, and Tim Johnson, 2018, Inflexibility and Stock Returns, *The Review of Financial Studies* 31, 278–321.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, 2015, The Value of Corporate Culture, *Journal of Financial Economics* 117, 60–76.

Hacamo, Isaac, 2023, Racial Prejudice in the Workplace and Sales Boycotts, *Available at SSRN 4033827*.

Hackbarth, Dirk, and Timothy Johnson, 2015, Real Options and Risk Dynamics, *The Review of Economic Studies* 82, 1449–1482.

Hambrick, Donald C., Theresa Seung Cho, and Ming-Jer Chen, 1996, The Influence of Top Management Team Heterogeneity on Firms' Competitive Moves, *Administrative science quarterly*, 659–684.

Harford, Jarrad, Qiyang He, and Buhui Qiu, 2023, Firm-Level Labor-Shortage Exposure, *Available at SSRN 4410126*.

Hoffman, L. R., and N. R. F. Maier, 1961, Quality and acceptance of problem solutions by members of homogeneous and heterogeneous groups., *The Journal of Abnormal and Social Psychology* 62, 401–407.

Hoffman, L. Richard, 1959, Homogeneity of Member Personality and Its Effect on Group Problem-Solving., *The Journal of Abnormal and Social Psychology* 58, 27.

Holmes, Oscar, Kaifeng Jiang, Derek R. Avery, Patrick F. McKay, In-Sue Oh, and C. Justice Tillman, 2021, A Meta-Analysis Integrating 25 Years of Diversity Climate Research, *Journal of Management* 47, 1357–1382.

Ibarra, Herminia, 1993, Personal Networks of Women and Minorities in Management: A Conceptual Framework, *Academy of management Review* 18, 56–87.

Ibarra, Herminia, 1995, Race, Opportunity, and Diversity of Social Circles in Managerial Networks, *Academy of management journal* 38, 673–703.

Ibarra, Herminia, 1997, Paving an Alternative Route: Gender Differences in Managerial Networks, *Social psychology quarterly*, 91–102.

Jackson, Susan E., 1992, Consequences of Group Composition for the Interpersonal Dynamics of Strategic Issue Processing, *Advances in Strategic Management* 8, 345–382.

Kets, Willemien, and Alvaro Sandroni, 2021, A Theory of Strategic Uncertainty and Cultural Diversity, *The Review of Economic Studies* 88, 287–333.

Klein, Katherine J., and Steve W. J. Kozlowski, 2000, From Micro to Meso: Critical Steps in Conceptualizing and Conducting Multilevel Research, *Organizational Research Methods* 3, 211–236.

Kölbel, Julian F, Markus Leippold, Jordy Rillaerts, and Qian Wang, 2022, Ask BERT: How Regulatory Disclosure of Transition and Physical Climate Risks Affects the CDS Term Structure*, *Journal of Financial Econometrics*, nbac027.

Koren, Miklós, and Rita Peto, 2020, Business Disruptions from Social Distancing, *PLOS ONE* 15.

Lang, Kevin, 1986, A Language Theory of Discrimination, *The Quarterly Journal of Economics* 101, 363–382.

Lang, Kevin, and Ariella Kahn-Lang Spitzer, 2020, Race Discrimination: An Economic Perspective, *Journal of Economic Perspectives* 34, 68–89.

Lu, Yan, Narayan Y. Naik, and Melvyn Teo, 2024, Diverse Hedge Funds, *The Review of Financial Studies* 37, 639–683.

Madera, Juan M., 2019, Top Management Gender Diversity and Organizational Attraction: When and Why It Matters., *Archives of Scientific Psychology* 7, 90.

Mannix, Elizabeth, and Margaret A. Neale, 2005, What Differences Make a Difference? The Promise and Reality of Diverse Teams in Organizations, *Psychological science in the public interest* 6, 31–55.

Marinescu, Ioana, Andrew Chamberlain, Morgan Smart, and Nadav Klein, 2021, Incentives Can Reduce Bias in Online Employer Reviews, *Journal of Experimental Psychology: Applied* 27, 393–407.

Marinescu, Ioana, Nadav Klein, Andrew Chamberlain, and Morgan Smart, 2018, Incentives Can Reduce Bias in Online Reviews, National Bureau of Economic Research.

Marquis, Christopher, Michael W. Toffel, and Yanhua Zhou, 2016, Scrutiny, Norms, and Selective Disclosure: A Global Study of Greenwashing, *Organization Science* 27, 483–504.

Matsa, David A., 2018, Capital structure and a firm's workforce, *Annual Review of Financial Economics* 10, 387–412.

Milliken, Frances J., and Luis L. Martins, 1996, Searching for Common Threads: Understanding the Multiple Effects of Diversity in Organizational Groups, *The Academy of Management Review* 21, 402.

Nemeth, Charlan J., 1986, Differential Contributions of Majority and Minority Influence., *Psychological review* 93, 23.

Nishii, Lisa H., 2013, The benefits of climate for inclusion for gender-diverse groups, *Academy of Management journal* 56, 1754–1774.

Phelps, Edmund S., 1972, The Statistical Theory of Racism and Sexism, *The American Economic Review* 62, 659–661.

Prasad, Anshuman, 2001, Understanding Workplace Empowerment as Inclusion: A Historical Investigation of the Discourse of Difference in the United States, *The Journal of Applied Behavioral Science* 37, 51–69.

Qiu, Yue, and Tracy Yue Wang, 2021, Skilled Labor Risk and Corporate Policies, *The Review of Corporate Finance Studies* 10, 437–472.

Rajan, Raghuram, Pietro Ramella, and Luigi Zingales, 2023, What Purpose Do Corporations Purport? Evidence from Letters to Shareholders, National Bureau of Economic Research.

Reinartz, Sebastian J., and Thomas Schmid, 2016, Production Flexibility, Product Markets, and Capital Structure Decisions, *The Review of Financial Studies* 29, 1501–1548.

Riordan, Christine M., 2014, Diversity Is Useless Without Inclusivity, *Harvard Business Review* 5.

Roberson, Quinetta M., 2006, Disentangling the Meanings of Diversity and Inclusion in Organizations, *Group & Organization Management* 31, 212–236.

Rubinstein, Yona, and Dror Brenner, 2014, Pride and Prejudice: Using Ethnic-Sounding Names and Inter-Ethnic Marriages to Identify Labour Market Discrimination, *Review of Economic Studies* 81, 389–425.

Sanh, Victor, Lysandre Debut, Julien Chaumond, and Thomas Wolf, 2019, DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter, *arXiv preprint arXiv:1910.01108*.

Serfling, Matthew, 2016, Firing Costs and Capital Structure Decisions, *The Journal of Finance* 71, 2239–2286.

Sherbin, Laura, and Ripa Rashid, 2017, Diversity Doesn't Stick Without Inclusion, *Harvard Business Review* 1, 2017.

Simintzi, Elena, Vikrant Vig, and Paolo Volpin, 2015, Labor Protection and Leverage, *The Review of Financial Studies* 28, 561–591.

Table 1: Descriptive statistics.

This table presents descriptive statistics for the main variables in my sample at the firm-year level. Panel A shows the summary statistics. Panel B shows the pairwise correlations among select variables. Detailed variable definitions are in Appendix B.

Panel A: Summary statistics

	N	Mean	1st Perc.	p25	Median	p75	99th Perc.
Rating - D&I	24668	3.438	1.333	3.036	3.48	3.855	5
Rating - overall	24668	3.218	1.5	2.836	3.229	3.623	4.978
Rating - balance	24624	3.289	1.667	2.917	3.3	3.676	5
Rating - culture	18869	3.217	1.571	2.784	3.222	3.65	4.833
Rating - career	24624	3.019	1.5	2.655	3	3.376	4.627
Rating - compensation	24624	3.274	1.75	2.9	3.296	3.667	4.6
Rating - management	24621	2.88	1	2.478	2.859	3.257	4.75
Size (log assets)	21720	7.946	3.524	6.641	7.865	9.187	12.787
ROA	20856	.118	-.364	.064	.117	.175	.466
Sales growth	21105	.064	-.489	-.021	.052	.135	.76
Tobin's q	19301	2.079	.733	1.129	1.547	2.355	9.521
Total debt/assets	21720	.268	0	.074	.235	.401	.953
cash/assets	21677	.16	.001	.035	.097	.227	.738
Refinitiv exec gender diversity	7355	15.407	0	0	14.29	25	50
Refinitiv board gender diversity	7355	20.577	0	12.5	20	27.27	50
Refinitiv board D&I policy	7357	.847	0	1	1	1	1
Refinitiv D&I controversy score	5255	98.815	75	100	100	100	100
Refinitiv D&I score	5255	18.934	0	8.5	13	26.5	63.5

Panel B: Pairwise correlations

D&I and firm characteristics

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Rating - D&I	1.000						
(2) Size (log asse~)	0.094***	1.000					
(3) ROA	-0.019**	0.062***	1.000				
(4) Sales growth	0.072***	-0.119***	0.078***	1.000			
(5) Tobin's q	0.185***	-0.224***	0.168***	0.290***	1.000		
(6) Total debt/ass~s	-0.090***	0.173***	0.053***	-0.091***	-0.097***	1.000	
(7) cash/assets	0.137***	-0.352***	-0.149***	0.175***	0.476***	-0.349***	1.000

*** p<0.01, ** p<0.05, * p<0.1*

D&I and diversity indicators from Refinitiv

Variables	(1)	(2)	(3)	(4)	(5)	(6)
(1) Rating - D&I	1.000					
(2) Refinitiv exec~s	0.139***	1.000				
(3) Refinitiv boar~r	0.186***	0.359***	1.000			
(4) Refinitiv boar~y	0.060***	0.073***	0.113***	1.000		
(5) Refinitiv D&I ~c	-0.077***	-0.070***	-0.073***	0.022	1.000	
(6) Refinitiv D&I ~e	0.254***	0.350***	0.412***	0.095***	-0.192***	1.000

*** p<0.01, ** p<0.05, * p<0.1*

D&I and other Glassdoor ratings

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Rating - D&I	1.000						
(2) Rating - overall	0.868***	1.000					
(3) Rating - balance	0.666***	0.687***	1.000				
(4) Rating - culture	0.907***	0.880***	0.769***	1.000			
(5) Rating - career	0.743***	0.829***	0.571***	0.806***	1.000		
(6) Rating - compe~n	0.556***	0.679***	0.522***	0.657***	0.654***	1.000	
(7) Rating - manag~t	0.807***	0.851***	0.674***	0.877***	0.788***	0.602***	1.000

*** p<0.01, ** p<0.05, * p<0.1*

Table 2: Validating the D&I measure.

This table presents tests to validate my D&I measure at the firm-year level, which is the average predicted D&I rating across reviews in each firm-year. Panel A shows coefficient estimates in logit models where the dependent variable is an indicator of whether a firm is in Fortune's Best Workplaces for Diversity list (Best Diversity list) in a year. Panel B shows estimates from a similar model predicting a firm's likelihood of landing in Forbes Magazine's Best Diversity List in 2023 based on firm characteristics in 2019. Panel C shows coefficient estimates in logit models where the dependent variable is an indicator of whether a firm is in DiversityInc's Top 50 list (DiversityInc list) in a year. Panel D shows how persistent the D&I measure is by regressing it on its lagged value and other control variables. Overall rating, culture rating, balance rating, compensation rating, career rating, and management rating are all the other ratings available on Glassdoor, averaged to the firm-year level across reviews, all measured before the outcome of prediction. The Refinitiv D&I rating is the average of diversity and inclusion scores provided by Refinitiv. Control variables include size, ROA, and leverage (total debts over assets). Detailed variable definitions are in Appendix B. Industry classifications are based on Fama-French 48 industries. All models include an intercept. Standard errors are clustered at the firm level.

Panel A: Predicting Fortune's Best Diversity List.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D&I rating	2.046***	2.338***	1.987***	3.003***	2.423***	2.852***	2.737***	1.637**
	(.176)	(.246)	(.233)	(.352)	(.368)	(.43)	(.624)	(.646)
Lagged Best Diversity			5.284***		4.776***			
			(.355)		(.351)			
Refinitiv D&I						.051***		
						(.008)		
Overall rating							-.416	
							(.576)	
Culture rating								.623
								(.607)
Observations	9780	5417	4325	5148	4109	1871	5417	5416
Pseudo R^2	.109	.156	.495	.273	.527	.254	.157	.157
Industry FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	Yes	Yes	No	No	No

Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Panel B: Predicting Forbes' Top 100 Firms on Diversity in 2023

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D&I rating$_{2019}$	1.648***	2.066***	1.863***	2.228***	2.001***	2.14***	2.801***	2.818***
	(.203)	(.338)	(.334)	(.5)	(.504)	(.526)	(.686)	(.772)
Fortune's Best Diversity$_{2019}$			1.756***		1.011**			
			(.379)		(.496)			
Refinitiv D&I$_{2019}$						.085***		
						(.01)		
Overall rating$_{2019}$							-.763	
							(.596)	
Culture rating$_{2019}$								-.662
								(.607)
Observations	1971	1181	1181	1151	1151	790	1181	1181
Pseudo R^2	.072	.159	.188	.403	.41	.405	.16	.16
Industry FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	Yes	Yes	No	No	No

Standard errors are in parentheses
*** p<.01, ** p<.05, * p<.1*

Panel C: Predicting DiversityInc's Top 50 list.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D&I rating	.68***	.838***	.442***	.695***	.157	.591	.663***	1.581***
	(.085)	(.104)	(.131)	(.213)	(.229)	(.468)	(.176)	(.225)
Lagged DiversityInc			7.445***		6.357***			
			(.275)		(.256)			
Refinitiv D&I						.09***		
						(.009)		
Overall rating							.211	
							(.199)	
Culture rating								-.305
								(.235)
Observations	21154	13106	13106	12740	12740	2592	13106	8871
Pseudo R^2	.019	.122	.751	.402	.781	.38	.123	.131
Industry FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	Yes	Yes	No	No	No

Standard errors are in parentheses
*** p<.01, ** p<.05, * p<.1*

Panel D: Persistence in the D&I measure.

	(1)	(2)	(3)	(4)	(5)	(6)
	D&I in full sample			D&I in sample with above median number of reviews		
Lagged D&I	.44***	.404***	.04***	.714***	.691***	.068***
	(.01)	(.011)	(.005)	(.008)	(.01)	(.005)
Rating - overall			.397***			.422***
			(.028)			(.014)
Rating - balance			.059***			.014*
			(.011)			(.008)
Rating - culture			.369***			.462***
			(.016)			(.011)
Rating - career			.037***			.027**
			(.013)			(.011)
Rating - compensation			-.033***			-.072***
			(.008)			(.007)
Rating - management			.093***			.01
			(.014)			(.011)
Observations	21973	17202	13929	12055	9600	8812
R-squared	.229	.257	.903	.569	.589	.951
Industry FE	No	Yes	Yes	No	Yes	Yes
Year FE	No	Yes	Yes	No	Yes	Yes

Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Table 3: Flexibility and D&I in the cross-section.

This table tests whether D&I firms have lower or higher flexibility than other firms, by regressing a firm's flexibility (FLEX) measured over 2009-2020 on the average firm characteristics measured over the same period. FLEX is the ratio of a firm's volatility of sales over assets to the firm's range of operating cost margins (Gu et al. 2018). The regressors include the D&I rating, board diversity variables, such as gender mix and nationality mix, and other firm characteristics like size, leverage, and labor stock. Variable definitions are in Appendix B. Industry classifications are based on Fama-French 48 industries. All models include an intercept. Robust standard errors are in parentheses.

	(1) FLEX	(2) FLEX	(3) FLEX	(4) FLEX	(5) FLEX
D&I rating	-.691***	-.503***	-.382***	-.353***	-.36***
	(.104)	(.103)	(.119)	(.124)	(.126)
Size			.048	.034	.039
			(.029)	(.035)	(.041)
Leverage			.11	.083	.068
			(.216)	(.238)	(.245)
Cash/assets			-1.765***	-1.921***	-1.916***
			(.41)	(.432)	(.446)
Labor stock			7.625**	6.884*	8.207**
			(3.443)	(3.6)	(3.627)
Tobin's Q			.033	.036	.038
			(.033)	(.036)	(.036)
Fixed assets ratio			-1.467***	-1.268***	-1.295***
			(.361)	(.392)	(.385)
Inst. ownership (%)				.026	.014
				(.256)	(.255)
Inst. ownership HHI				.196	.116
				(.616)	(.623)
Inside ownership				-.007	.02
				(.047)	(.047)
Board nationality mix					.128
					(.345)
Board gender ratio					-.095
					(.588)
Board age diversity					-.066**
					(.026)
Board diversity: firm tenure					-.024*
					(.013)
Board diversity: qualifications					.032
					(.157)
Board: network size					0
					(0)
Observations	1813	1813	1746	1618	1605
R-squared	.021	.152	.175	.168	.176
Fixed effects	No	Industry	Industry	Industry	Industry

Table 4: Corporate flexibility after a shock to D&I.

This table investigates how a firm's D&I rating and flexibility changed after a plausibly exogenous shock to D&I: the 7th Circuit Court ruling in 2013. The table shows difference-in-difference tests in which the treated firms are firms headquartered in the states covered by the Seventh Circuit Court (Treat = 1) and the control firms are other US firms (Treat = 0). I compute flexibility (FLEX) as the ratio of a firm's volatility of sales over assets to the firm's range of operating cost relative to sales during a period (Gu et al. 2018). I do so for the two five-year periods before and after the court ruling: 2008-2012 and 2013-2017. The post indicator equals one for the later period, and zero for the former period. The D&I rating for the post-shock period is the average D&I rating across the years in that period. The D&I rating for the pre-shock period is simply the D&I rating in 2012, right before the court ruling. Columns (1) and (2) are for the full sample, whereas the next four columns are for the two different subsamples based on prior D&I rating: high (above median) and low (below median) D&I rating in 2012. Detailed variable definitions are in Appendix B. Standard errors are clustered at the state level.

	(1)	(2)	(3)	(4)	(5)	(6)
	Overall sample		**Low prior D&I subsample**		**High prior D&I subsample**	
	D&I rating	FLEX	D&I rating	FLEX	D&I rating	FLEX
Treat * Post	.049	-1.136***	.117***	-1.611**	-.047	.022
	(.041)	(.391)	(.026)	(.606)	(.071)	(.339)
_cons	3.363***	4.029***	2.994***	4.344***	3.722***	3.669***
	(.002)	(.018)	(.001)	(.029)	(.003)	(.015)
Observations	2952	2914	1454	1340	1498	1364
R-squared	.738	.641	.694	.595	.642	.696
Controls	No	No	No	No	No	No
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Table 5: Flexibility in responding to COVID shock.

This table examines whether D&I firms responded better to the COVID shock than other firms, by regressing the change in a firm's performance between 2019 and 2020 on an indicator of high exposure to the COVID shock and its interactions with the firm's D&I rating and other characteristics, all measured in 2019 (lagged one year). Performance is captured by return on assets (ROA). Firm characteristics other than the D&I rating are standardized to have a mean of zero and a standard deviation of one in the full sample: 2008-2020. Detailed variable definitions are in Appendix B. Industry classifications are based on Fama-French 48 industries. All models include an intercept and the standalone variables for each featured interaction term, but I do not show them for brevity. Robust standard errors are in parentheses.

Change in ROA (2019-2020)

	(1)	(2)	(3)	(4)	(5)	(6)
exposure	-.026***	.086***	.079***	.08***	.082***	
	(.004)	(.028)	(.028)	(.029)	(.029)	
exposure * lagged D&I rating		-.032***	-.028***	-.027***	-.028***	-.02**
		(.008)	(.008)	(.008)	(.009)	(.009)
exposure * lagged Leverage			-.009**	-.003	-.004	0
			(.004)	(.005)	(.005)	(.005)
exposure * lagged Cash/assets			-.004	-.009	-.009	-.005
			(.007)	(.008)	(.008)	(.007)
exposure * lagged Size			.01**	.01**	.008	.014**
			(.004)	(.005)	(.005)	(.006)
exposure * lagged Tobin's Q				.002	.002	.001
				(.005)	(.005)	(.005)
exposure * lagged Inst. ownership (%)				-.004	-.005	0
				(.005)	(.005)	(.005)
exposure * lagged Inside ownership				-.001	-.001	-.003
				(.004)	(.004)	(.004)
exposure * lagged Fixed assets ratio				-.009*	-.008*	-.002
				(.005)	(.005)	(.005)
exposure * lagged Advertising/sales					-.006	-.006
					(.005)	(.006)
exposure * lagged Labor stock					-.004	-.006
					(.008)	(.008)
exposure * lagged H1-B reliance					.011	.014
					(.008)	(.009)
Observations	1276	1229	1229	1077	1077	1069
R-squared	.037	.048	.1	.138	.145	.255
Fixed effects	No	No	No	No	No	Industry

Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Table 6: Workforce and operating efficiency during COVID shock.

This table examines whether D&I firms managed workforce and operating efficiency differently during the COVID shock compared to other firms, by regressing the change in a firm's employee count and efficiency measures between 2019 and 2020 on an indicator of high exposure to the COVID shock and its interactions with the firm's D&I rating and other characteristics, all measured in 2019 (lagged one year). The dependent variables are the change in employee count (Panel A), the change in selling, administrative and selling expense relative to sales (Panel B), the change in sales growth (Panel C), and the change in sales per employee, in thousands (Panel D). Detailed variable definitions are in Appendix B. All scaled measures are winsorized at the 1% and 99% percentile. Industry classifications are based on Fama-French 48 industries. All models include an intercept and the standalone variables for each featured interaction term, but I do not show them for brevity. Robust standard errors are in parentheses.

Panel A: Change in employee count.

	(1)	(2)	(3)	(4)	(5)	(6)
exposure	-.065***	.004	-.058	-.051	-.036	
	(.011)	(.081)	(.084)	(.087)	(.088)	
exposure * lagged D&I rating		-.019	.006	.005	.005	.015
		(.023)	(.023)	(.024)	(.024)	(.024)
exposure * lagged Leverage			-.003	-.009	-.013	.004
			(.012)	(.012)	(.012)	(.012)
exposure * lagged Cash/assets			.002	-.003	-.004	.006
			(.02)	(.02)	(.02)	(.019)
exposure * lagged Size			-.006	-.014	-.017	-.013
			(.013)	(.013)	(.014)	(.015)
exposure * lagged Tobin's Q				.015	.016	.013
				(.011)	(.011)	(.011)
exposure * lagged Inst. ownership (%)				-.039***	-.04***	-.038***
				(.015)	(.014)	(.014)
exposure * lagged Inside ownership				-.01	-.009	-.013
				(.016)	(.015)	(.013)
exposure * lagged Fixed assets ratio				-.004	-.005	.014
				(.013)	(.013)	(.017)
exposure * lagged Advertising/sales					.016	.028*
					(.014)	(.016)
exposure * lagged Labor stock					0	-.006
					(.022)	(.024)
exposure * lagged H1-B reliance					-.013	-.014
					(.017)	(.018)
Observations	1285	1239	1239	1078	1078	1071
R-squared	.028	.043	.089	.201	.21	.292
Fixed effects	No	No	No	No	No	Industry

Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Panel B: Change in SG&A expense relative to sales.

	(1)	(2)	(3)	(4)	(5)	(6)
exposure	.009***	-.023	-.011	-.022	-.016	
	(.003)	(.017)	(.019)	(.02)	(.02)	
exposure * lagged D&I rating		.009*	.005	.008	.006	.003
		(.005)	(.005)	(.006)	(.006)	(.006)
exposure * lagged Leverage			0	.001	.001	-.001
			(.003)	(.003)	(.003)	(.003)
exposure * lagged Cash/assets			.006	.014**	.012**	.008
			(.005)	(.006)	(.006)	(.006)
exposure * lagged Size			.004	.003	.003	-.002
			(.003)	(.004)	(.004)	(.004)
exposure * lagged Tobin's Q				-.005**	-.005**	-.005*
				(.003)	(.003)	(.003)
exposure * lagged Inst. ownership (%)				.001	.001	-.001
				(.003)	(.003)	(.003)
exposure * lagged Inside ownership				0	0	.001
				(.003)	(.003)	(.003)
exposure * lagged Fixed assets ratio				.008**	.008**	.002
				(.003)	(.003)	(.004)
exposure * lagged Advertising/sales					.002	0
					(.004)	(.004)
exposure * lagged Labor stock					.001	.004
					(.006)	(.006)
exposure * lagged H1-B reliance					.003	.002
					(.004)	(.005)
Observations	1197	1153	1153	1009	1009	1002
R-squared	.009	.011	.029	.05	.067	.136
Fixed effects	No	No	No	No	No	Industry

Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Panel C: Change in sales growth

	(1)	(2)	(3)	(4)	(5)	(6)
exposure	-.056***	.206**	.177**	.234**	.258***	
	(.012)	(.084)	(.085)	(.094)	(.095)	
exposure * lagged D&I rating		-.075***	-.063***	-.074***	-.08***	-.057**
		(.024)	(.024)	(.027)	(.027)	(.026)
exposure * lagged Leverage			-.003	.005	.001	.017
			(.013)	(.016)	(.016)	(.014)
exposure * lagged Cash/assets			0	-.008	-.011	.008
			(.017)	(.021)	(.021)	(.02)
exposure * lagged Size			-.006	-.019	-.018	-.003
			(.014)	(.017)	(.017)	(.017)
exposure * lagged Tobin's Q				.017	.016	.018
				(.011)	(.011)	(.011)
exposure * lagged Inst. ownership (%)				-.023	-.024	-.015
				(.016)	(.016)	(.015)
exposure * lagged Inside ownership				.008	.008	-.001
				(.013)	(.012)	(.012)
exposure * lagged Fixed assets ratio				-.01	-.009	.02
				(.015)	(.015)	(.019)
exposure * lagged Advertising/sales					.024	.03
					(.015)	(.019)
exposure * lagged Labor stock					.002	-.001
					(.024)	(.025)
exposure * lagged H1-B reliance					.021	.031*
					(.014)	(.018)
Observations	1294	1247	1247	1079	1079	1071
R-squared	.016	.023	.035	.066	.077	.216
Fixed effects	No	No	No	No	No	Industry

Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Panel D: Change in sales per employee.

	(1)	(2)	(3)	(4)	(5)	(6)
exposure	3.746	94.598**	110.719***	103.343**	101.783**	
	(4.958)	(39.176)	(40.462)	(41.378)	(41.545)	
exposure * lagged D&I rating		-26.22**	-29.634**	-26.528**	-26.999**	-20.657*
		(11.252)	(11.503)	(11.604)	(11.627)	(11.851)
exposure * lagged Leverage			-7.201	-2.986	-2.518	-5.423
			(5.929)	(7.115)	(7.036)	(7.48)
exposure * lagged Cash/assets			7.968	-1.614	-1.451	5.022
			(8.947)	(10.471)	(10.516)	(9.777)
exposure * lagged Size			18.636***	8.97	9.627	12.414*
			(6.384)	(7.233)	(7.753)	(7.375)
exposure * lagged Tobin's Q				7.831	7.474	4.399
				(6.023)	(6.138)	(6.113)
exposure * lagged Inst. ownership (%)				7.282	6.758	11.116
				(7.114)	(7.214)	(6.881)
exposure * lagged Inside ownership				8.789	8.179	10.7
				(11.064)	(10.944)	(10.497)
exposure * lagged Fixed assets ratio				.257	.91	8.283
				(6.206)	(6.213)	(6.617)
exposure * lagged Advertising/sales					-3.676	-5.33
					(6.809)	(6.726)
exposure * lagged Labor stock					-.807	.064
					(6.927)	(7.559)
exposure * lagged H1-B reliance					17.287*	23.469*
					(10.163)	(12.068)
Observations	1285	1239	1239	1078	1078	1071
R-squared	0	.006	.026	.044	.052	.221
Fixed effects	No	No	No	No	No	Industry

Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Figure 1: Graphs of D&I and Flexibility around the D&I-related court ruling

This figure plots the typical difference-in-differences graph around the circuit court ruling related to diversity and inclusion (D&I) issues in 2013. In particular, it plots the regression coefficients (along with the 95% confidence intervals) on the interactions between the treatment indicator (equaling one for firms headquartered in Indiana, Illinois, and Wisconsin, and zero otherwise) and year indicators relative to the treatment year: 2013. The indicator for the year t-1, or 2012, is omitted because 2012 is chosen as the reference year. The dependent variables include the D&I rating and the flexibility measure (FLEX). FLEX is measured over a five-year period (e.g., 2013-2017). See Appendix B for detailed variable descriptions. All regressions include a constant, firm fixed effects, and year fixed effects. The 95% confidence intervals are based on standard errors that are clustered at the state level. Figure 1a shows the results for the full sample, while figures 1b and 1c show the results for the subsamples with prior low (below median) D&I rating and prior high (above median) D&I rating measured in 2012, the year before the court ruling.

Figure 1a: Full sample.

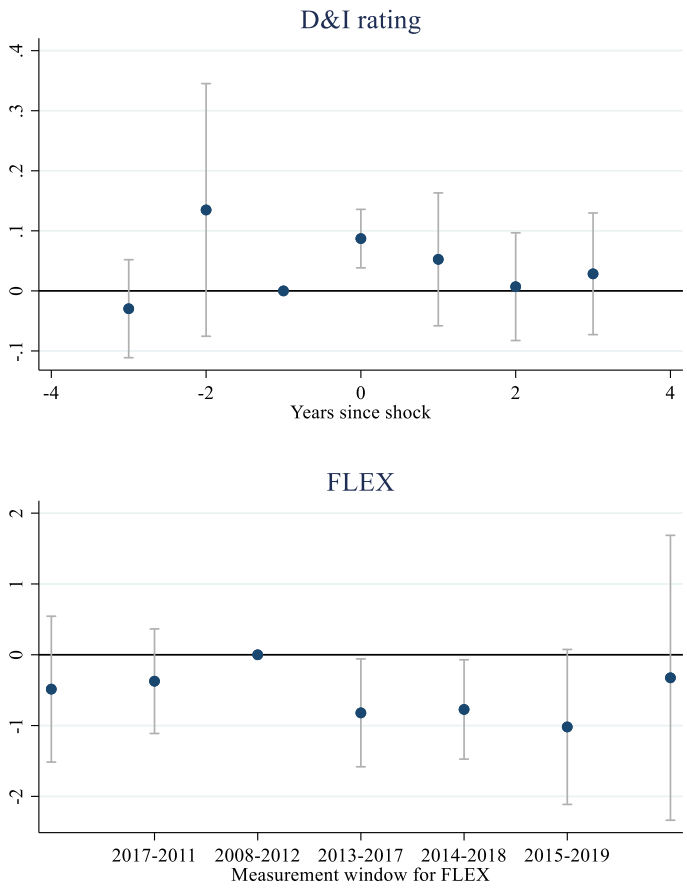


Figure 1b: Subsample – firms with prior low D&I rating.

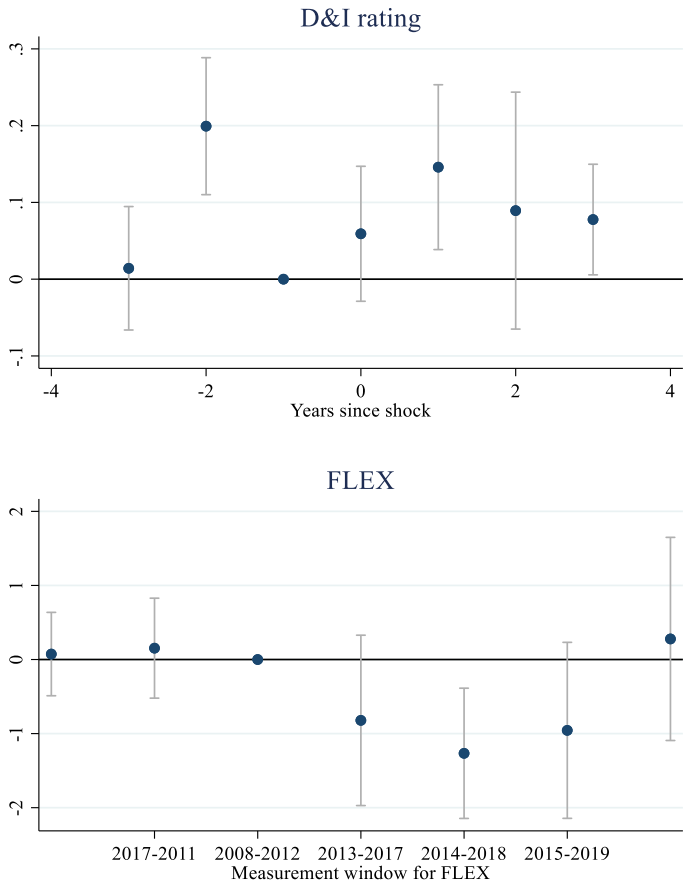
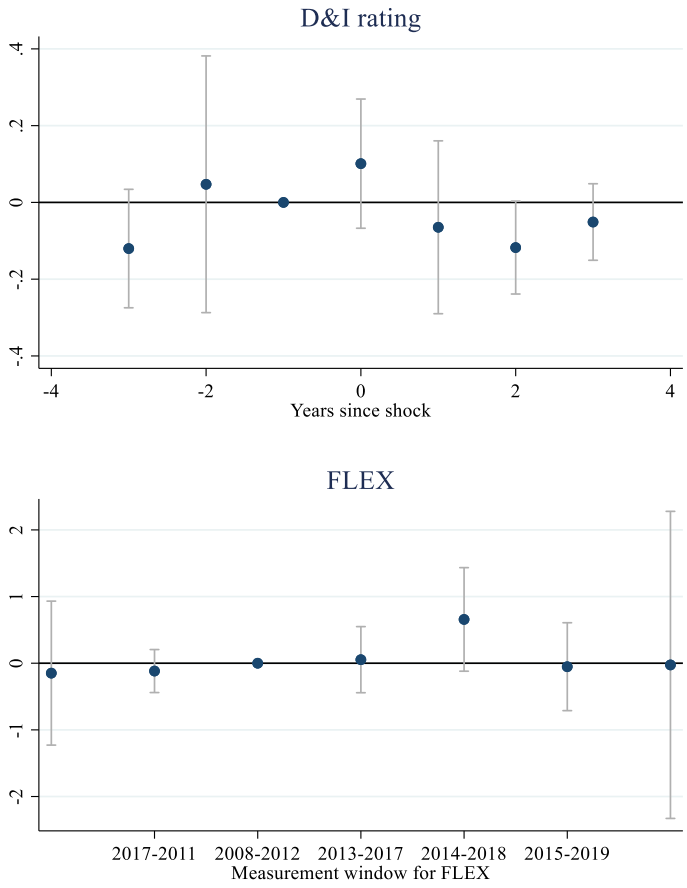


Figure 1c: Subsample – firms with prior high D&I rating.



ONLINE APPENDIX

Appendix A: Glassdoor review example

This appendix shows an example of a Glassdoor review (Panel A), and the combined text based on this review as an example of the text input I will feed into the BERT model to predict the D&I rating (Panel B). This specific review was written on April 30, 2021, by a human resources employee of Target Corporation, located in Homestead, PA.

Panel A: Example review

Culture over everything

Apr 30, 2021 - Human Resources Expert in Homestead, PA

✗ Recommend ✓ CEO Approval ✓ Business Outlook

Pros

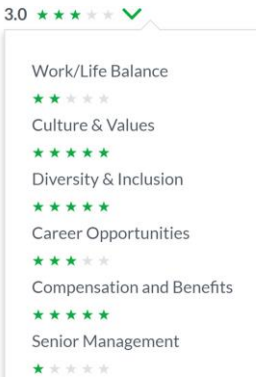
It is LGBTQIA+ friendly, they don't discriminate based on appearances (tattoos and colored hair are welcomed), livable wage, beautiful team culture

Cons

Upward mobility is very difficult within the same store and varies based on position, upper management/leadership really needs better training, high turnover with employees due to extreme expectations placed on team members

Advice to Management

Listen to your team. Build a genuine connection. Target refers to its team members as just that, team members. Treat them as such.



Panel B: Combined text as input to the BERT model.

"Pros are as follows: It is LGBTQIA+ friendly, they don't discriminate based on appearances (tattoos and colored hair are welcomed), livable wage, beautiful team culture. Cons are as follows: Upward mobility is very difficult within the same store and varies based on position, upper management/leadership really needs better training, high turnover with employees due to extreme expectations placed on team members. So overall, Culture over everything. I give an overall rating of 3.0, a work/life balance rating of 2.0, a culture and value rating of 5.0, a career opportunity rating of 3.0, a compensation and benefit rating of 5.0, and a senior management rating of 1.0."

Appendix B: Variable definitions

Below I list the detailed description of the variables used in my analyses. All variables are at the firm-year level, unless stated otherwise.

Rating – D&I, or **D&I rating**, or simply **D&I:** the average D&I rating across reviews in a firm-year, where the D&I rating for each review is the extrapolated D&I rating based on a BERT model on review texts and other ratings, unless the actual D&I rating is available. Source: Glassdoor and my own model.

Industry D&I: the average of D&I rating across reviews in firms outside of my sample that belongs to a specific industry. Source: Glassdoor.

Rating – overall, or **overall rating:** the average overall rating across reviews in a firm-year. Source: Glassdoor.

Rating – balance: the average rating of work-life balance across reviews in a firm-year. Source: Glassdoor.

Rating – culture, or **culture rating:** the average rating of culture and values across reviews in a firm-year. Source: Glassdoor.

Rating – career: the average rating of career path across reviews in a firm-year. Source: Glassdoor.

Rating – compensation, or **compensation rating:** the average rating of compensation and benetis across reviews in a firm-year. Source: Glassdoor.

Rating – management: the average rating of senior management across reviews in a firm-year. Source: Glassdoor.

Refinitiv D&I: the average of the diversity score and the inclusion score from Refinitiv for each firm-year. Source: Refinitiv Assets4.

Refinitiv exec gender diversity: the percentage of executives that are female. Source: Refinitiv Assets4.

Refinitiv board gender diversity: the percentage of directors that are female in a firm's board. Source: Refinitiv Assets4.

Refinitiv board D&I policy: an indicator of whether a firm has a pro-D&I policy regarding its board. Source: Refinitiv Assets4.

Refinitiv D&I controversy: the score Refinitiv assigns to a firm based on its controversies on D&I. Source: Refinitiv Assets4.

FLEX: The inverse of INFLEX measure in Gu et al. (2018); where INFLEX is calculated as ratio of a firm's range of operating cost margins over the standard deviation of the firm's sales over assets during a period.

Size (log assets): the logarithm of total assets. Source: Compustat.

ROA: Return on assets, or the ratio of earnings before interests, taxes, and depreciation to lagged total assets. Source: Compustat.

Sales growth: the log of the ratio of sales in year t to sales in year t-1. Source: Compustat.

Tobin's Q: (total assets - book equity - current liabilities + market equity)/total assets. Source: Compustat.

Cash/assets: cash and cash equivalents divided by total assets. Source: Compustat.

Total debt/assets, or **Total book leverage**, or **Leverage:** the sum of total long-term debt and total current debt, divided by total assets. Source: Compustat.

Capex/Assets: the ratio of capital expenditures (capx) to total assets (at), set to zero if missing capx. Source: Compustat.

Entrenchment index: A measure of corporate governance quality, constructed as in Bebchuk, Cohen, and Ferrell (2009). It is based on six provisions on corporate governance: staggered boards, limits to shareholder bylaw amendments, poison pills, golden parachutes, and supermajority requirements for mergers and charter amendments. Source: ISS Risk Metrics.

Fortune's Best Diversity: an indicator equaling one for firms that get in Fortune's list of Best Workplaces for Diversity in a year. Source: Fortune magazine.

Forbes' Best Diversity: an indicator equaling one for firms that get in Forbes Magazine's list of 100 Best Employers for Diversity in a year. Source: Forbes magazine.

DiversityInc's Top 50 list: an indicator equaling one for firms that get in the top 50 by DiversityInc in a year. Source: DiversityInc website.

Labor stock: the ratio of employee count to total assets. Source: Compustat.

Labor skill, or H1-B reliance: the ratio of a firm's number of H1-B visa applications per year, scaled by the firm's employee count in that year. In the COVID analyses, **H1-B reliance** refers to this ratio's average value over the 2010-2016 period for each firm. The data are from Qiu and Wang (2021).

Fixed assets ratio: the ratio of net Property Plant and Equipment to total assets. Source: Compustat.

Inst. ownership (%): the sum of dollar value of institutional ownership, divided by the sum of market value across securities per firm-year. Source: Thomson Reuters 13F.

Inst. ownership HHI: the sum of squared ownership weights of different institutional investors in a firm' stock. Source: Thomson Reuters 13F.

Inside ownership: the total equity-linked wealth of all directors in a firm's board and executive team, divided by the total market value of a firm's outstanding shares. Source: BoardEx.

Board nationality mix: the percentage of a firm's board that is from different countries. Source: BoardEx.

Board gender ratio: the ratio of males to females on a firm's board. Source: BoardEx.

Board age diversity: the standard deviation of age across directors in a firm's board. Source: BoardEx.

Board diversity: firm tenure: the standard deviation of years working at a firm across directors in the firm's board. Source: BoardEx.

Board: network size: the average network size of directors in a firm's board, where network size captures the number of overlaps through employment, other activities, and education with other directors in the database. Source: BoardEx.

Board diversity: qualifications: the standard deviation of the number of qualifications a director holds, across directors in a firm's board. Source: BoardEx.

SG&A/sales: the ratio of SG&A expenses to sales. Source: Compustat.

Exposure: an indicator of high exposure (above-sample-median) to the COVID shock, where the raw exposure of a firm is based on an industry score of how much the firm's industry relies on in-person interactions (Koren and Peto (2020)). Source: Koren and Peto (2020).

Advertising/sales: the ratio of advertising expenses to sales, set to zero if missing advertising expenses. Source: Compustat.

Institutional block ownership: an indicator equaling one if a firm has an institutional investor that is a block investor, i.e., holding at least 5% of the firm's shares outstanding. Source: Thomson Reuters 13F.

Sales/employees: the ratio of sales to number of employees (thousand dollars per employee). Source: Compustat.

Sale/assets, or **assets turnover**: the ratio of sales to total assets. Source: Compustat.

IA1: Sample construction

To construct my main sample, I start with the list of employers on Glassdoor as of July 2020, each represented by an ID created by Glassdoor (Glassdoor ID). First, I retain the Glassdoor IDs pertaining to companies or subsidiaries (i.e., removing nonprofits or governmental organizations) on Glassdoor with at least 100 reviews as of July 2020. Then I remove the companies or subsidiaries with a foreign headquarter, e.g., Mumbai (India), arriving at 7,851 unique Glassdoor IDs.

Next, I match each Glassdoor ID to a firm in the Compustat database for the 2008-2020 period. The matching is primarily based on stock ticker, which Glassdoor provides for 1,805 Glassdoor IDs. When a ticker is not available or not unique, I conduct fuzzy matching based on company names using the *matchit* package in Stata, after removing common string patterns, such as *Inc., Co., Limited, …* from the company names. I keep only the matches with a similarity score of 80% or above. When the similarity score is below one, i.e., the match is not exact (291 cases), I manually verify each match. Among the remaining unmatched Glassdoor IDs, 370 cases belong to a public company or a public company's subsidiary. In these cases, I manually verify and match each Glassdoor ID with a company in the Compustat database for the 2008-2020 period.[12] In the end, the final sample includes 2,199 distinct Glassdoor IDs linked to 2,113 firms that were ever publicly listed between 2008 and 2020.

[12] For each Glassdoor ID, I look up the corresponding company name on Google or the CRSP database to find if the company was ever publicly listed during 2008-2020. If not, I consider the company to be privately held and thus record "NA" for the company's CUSIP. If the firm was ever publicly listed during 2008-2020, I find the corresponding latest CUSIP for the company' stock. If the company changed its CUSIP during the sample period, I keep the latest CUSIP unless the change was very recent (2019 onwards). If the company was acquired by or merged with another company between 2008 and 2018, I drop the company from my sample. However, if the company was acquired or merged with another in 2019 or later, I keep the company's latest CUSIP as of 2018. If a company was a subsidiary of a publicly listed firm during the entire sample period, I assign the CUSIP of the parent firm to the subsidiary. Finally, I drop those Glassdoor IDs associated with a subsidiary of a foreign corporation from my sample.

Figure IA1: Distribution of the D&I rating at the firm-year level.

This figure shows the distribution of the D&I rating at the firm-year level for my sample, and compares that with the distribution of other ratings on Glassdoor.

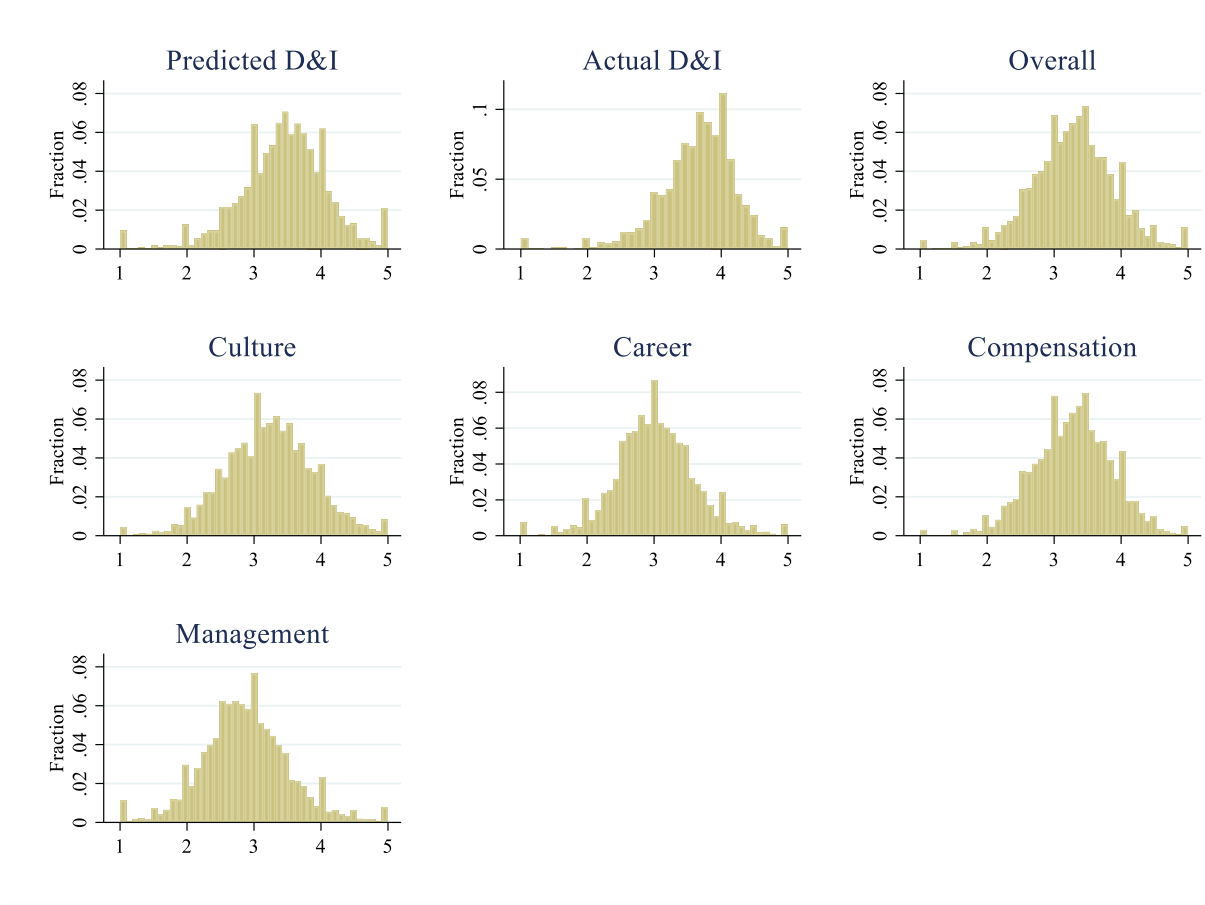


Figure IA2: Trends in D&I

This figure shows the evolution of my D&I measure for the average firm in my sample over 2009-2020, and compares that with the Refinitiv rating of D&I policies. The years 2008 and 2009 are omitted due to the limited coverage of Glassdoor data in its inception years. Refinitiv data are only available since 2016.

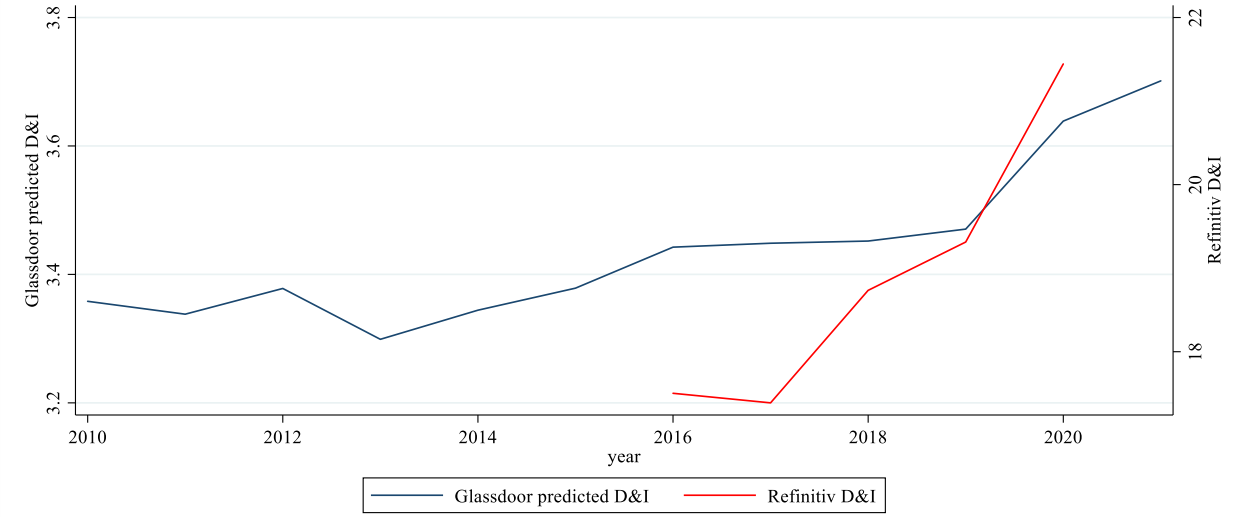


Figure IA3: D&I and Flexibility around the court ruling – matched sample

This figure replicates Figure 1 – Diff-in-diff graphs on a matched sample. It plots the regression coefficients) on the interactions between the treatment indicator (equaling one for firms headquartered in Indiana, Illinois, and Wisconsin, and zero otherwise) and year indicators relative to the treatment year: 2013. The indicator for the year t-1, or 2012, is omitted because 2012 is chosen as the reference year. The dependent variables include the D&I rating and the flexibility measure (FLEX). FLEX is measured over a five-year period (e.g., 2013-2017). All regressions include a constant, firm fixed effects, and year fixed effects. The 95% confidence intervals are based on standard errors that are clustered at the state level. The regressions are performed on a matched sample, in which each treated firm is matched to up to three control firms based on a nearest neighbor matching based on a firm's size and D&I ratings in 2011 and 2012, within the same Fama-French 48 industry. Figure 1a shows the results for the whole matched sample, while figures 1b and 1c show the results for the subsamples with prior low (below median) D&I rating and prior high (above median) D&I rating measured in 2012, the year before the court ruling.

Figure 1a: Full sample.

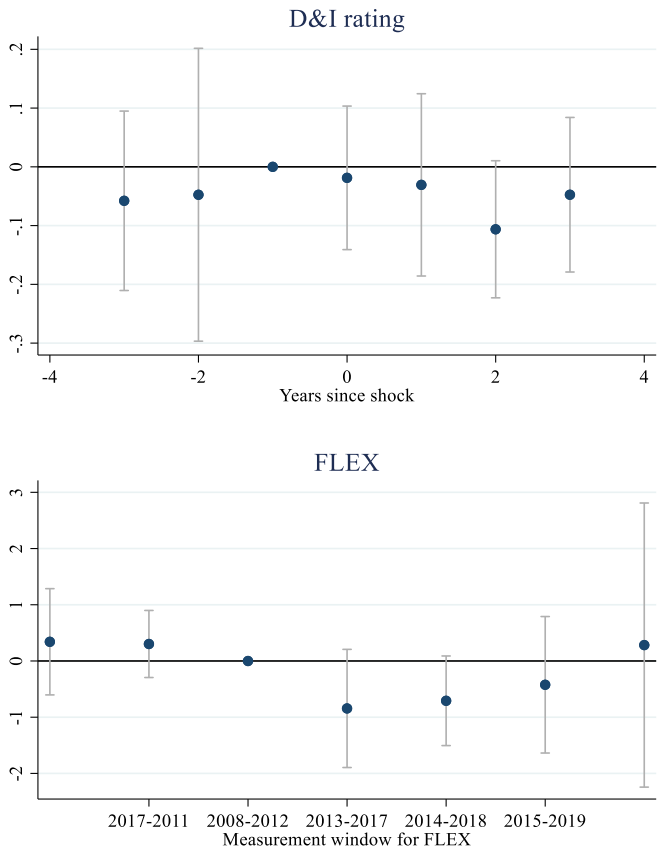


Figure 1b: Subsample – firms with prior low D&I rating.

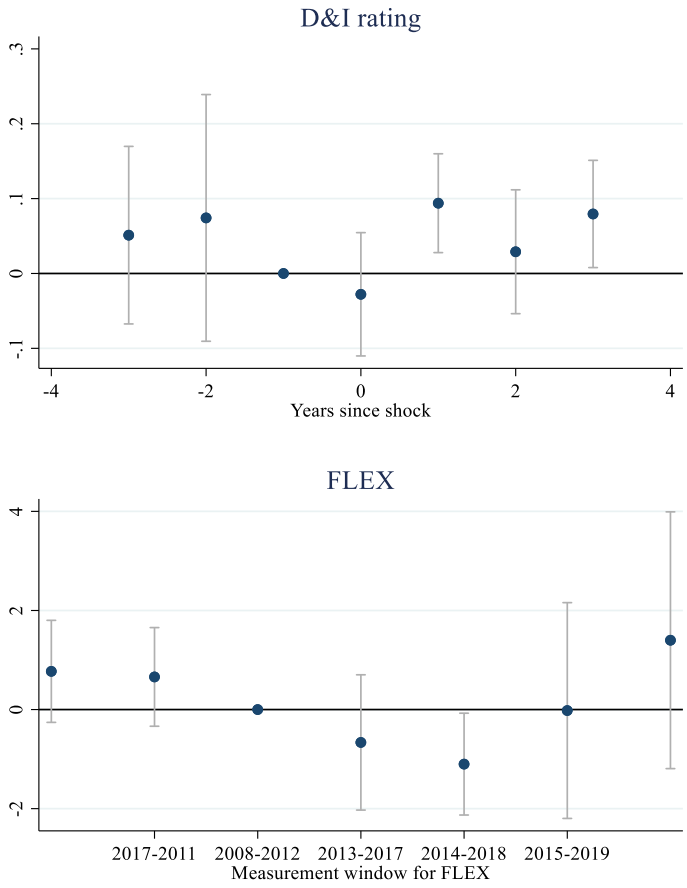


Figure 1c: Subsample – firms with prior high D&I rating.

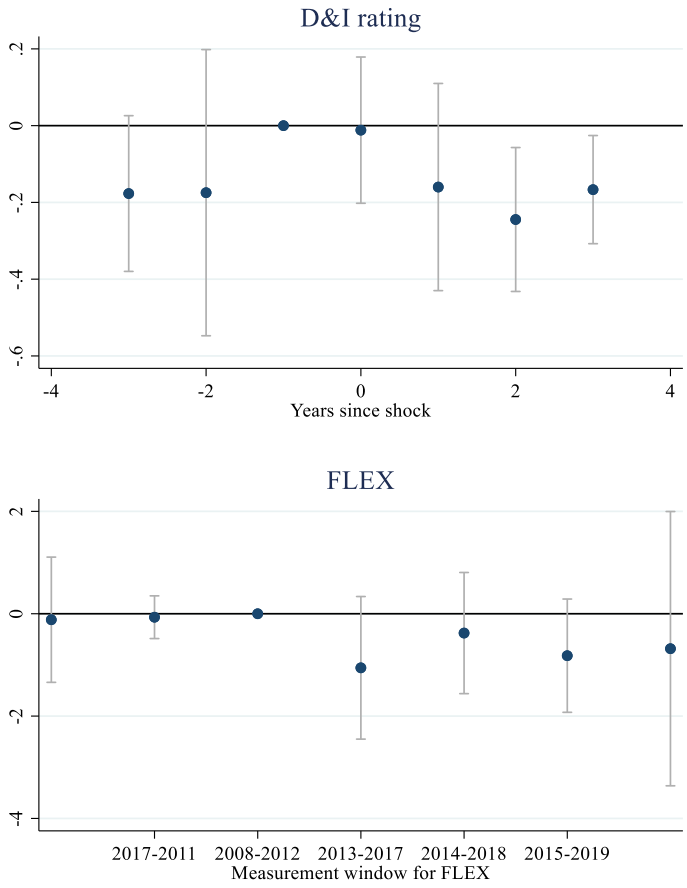


Table IA1: Summary statistics at the review level

This appendix shows the summary statistics of 10.4 million reviews (Panel A), the correlations among Glassdoor ratings at the review level (Panel B), and the training statistics of the BERT models to predict the D&I rating (Panel C), and the distribution of the actual vs. predicted D&I ratings at the review level (Panel D).

Panel A: Summary statistics of all reviews

	N	Mean	1st Perc.	p25	Median	p75	99th Perc.
Pros length	10425401	20.139	5	6	11	22	141
Cons length	10425401	28.947	5	7	12	29	254
Title length	10425401	4.208	1	2	3	5	17
Rating - overall	10425401	3.425	1	2	4	5	5
Rating - D&I	1194918	3.739	1	3	4	5	5
Rating - balance	8959847	3.331	1	2	3.5	5	5
Rating - culture	8484406	3.38	1	2	4	5	5
Rating - career	8964173	3.196	1	2	3	4	5
Rating - compensation	8952061	3.265	1	2	3	4	5
Rating - management	8853230	3.031	1	2	3	4	5
All other ratings available	10425401	.797	0	1	1	1	1

Panel B: Pairwise correlation across ratings

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Rating - overall	1.000						
(2) Rating - balance	0.680	1.000					
(3) Rating - culture	0.823	0.664	1.000				
(4) Rating - career	0.787	0.564	0.723	1.000			
(5) Rating - compensation	0.670	0.533	0.606	0.654	1.000		
(6) Rating - management	0.818	0.657	0.803	0.736	0.609	1.000	
(7) Rating - D&I	0.699	0.594	0.761	0.635	0.555	0.705	1.000

Panel C: Training sample statistics

Model	**Simpler**	**Fuller**
Input	**Title, Pros, Cons, Overall rating**	**Title, Pros, Cons, Overall rating, Balance rating, Culture rating, Management rating, Compensation rating**
Accuracy (% predicted = actual)	**54.549%**	**63.834%**

Panel D: Distribution of the predicted and actual D&I ratings

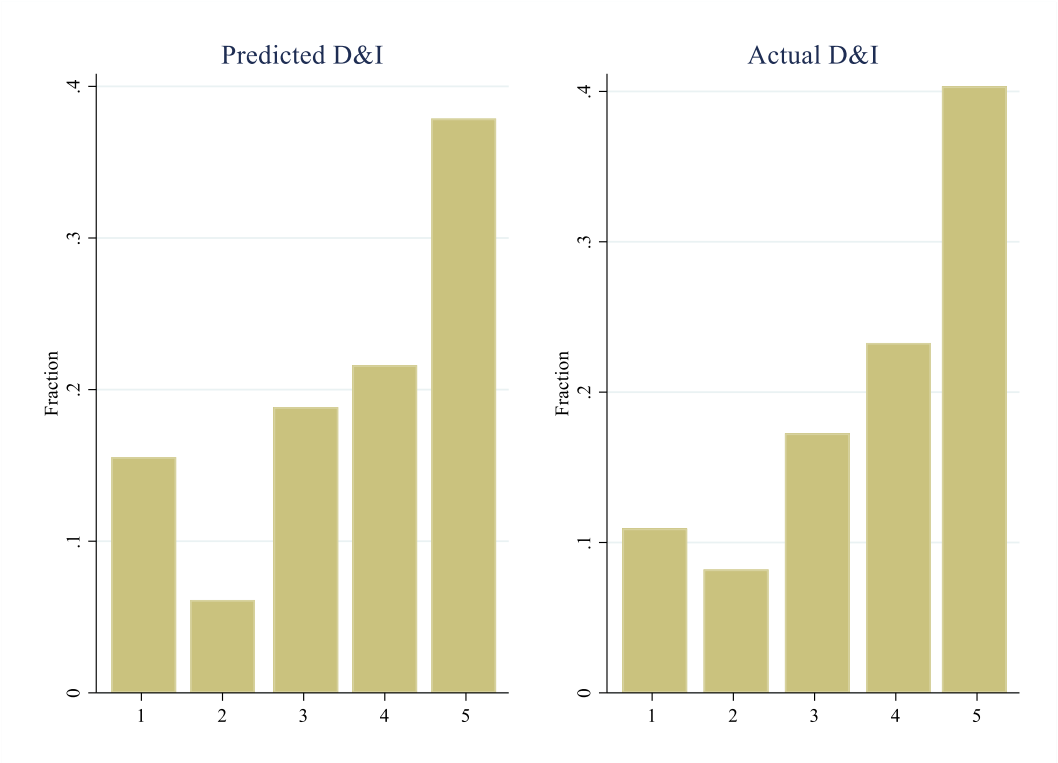


Table IA2: Words most strongly associated with the D&I rating.

The first column of this table lists the words that are mentioned most frequently in the pros section of reviews with a high estimated D&I rating (4 or 5 stars) relative to the pros section in reviews with a low estimated D&I rating (1 or 2 stars). The second column, by contrast, focuses on words in the cons section that are mentioned most frequently among reviews with a *low* D&I rating relative to reviews with a *high* D&I rating. I follow Acikalin et al. (2022)'s methodology to construct this table (see their Table 4), except the filtering criteria that I keep only words appearing in at least 0.1% of the reviews.[13]

Top 30 words in the pros section	Top 30 words in the cons section
inclusive, listens, transparent, encourages, invests, suggestions, diversity, transparency, encouraging, diverse, collaborative, feedback, cares, mentorship, values, approachable, culture, professionally, encouraged, voice, rewarded, wellbeing, promotes, excellence, grown, continuous, input, opinions, emphasis, coaching	racist, discrimination, sexual, bullying, harassment, backstabbing, unethical, lies, bully, male, illegal, boys, diversity, lie, lied, white, dishonest, fake, lying, women, hostile, abusive, woman, toxic, men, female, inappropriate, himself, rampant, nepotism

[13] Following the same filter as in Acikalin et al. is too restrictive as it would cut my corpus from 55,000 words to only 267 words. The results remain similar when I keep only words appearing in at least 0.01% of the reviews.

Table IA3: Top and bottom firms by D&I

This figure shows the top ten and bottom ten firms ranked based on the average D&I measure during 2016-2020 for the biggest 500 firms in my sample (based on their average total assets during the same time period).

Company name	Rating - D&I
SOUTHWEST AIRLINES	4.475
FACEBOOK INC	4.443
ALPHABET INC	4.437
SALESFORCE.COM INC	4.435
NVIDIA CORP	4.416
AVALONBAY COMMUNITIES INC	4.363
DELTA AIR LINES INC	4.345
INTL BANCSHARES CORP	4.333
ADOBE INC	4.326
APPLE INC	4.313
SS&C TECHNOLOGIES HLDGS INC	2.790
VALLEY NATIONAL BANCORP	2.778
ALEXION PHARMACEUTICALS INC	2.771
STERLING BANCORP	2.751
GRUPO TELEVISA SAB	2.733
FLEETCOR TECHNOLOGIES INC	2.727
NORFOLK SOUTHERN CORP	2.691
PACIFICORP	2.682
CSX CORP	2.651
UNION PACIFIC CORP	2.287

Table IA4: Determinants of D&I at the firm-year level

This table shows the estimates of regressing the firm-year D&I rating on different fixed effects and firm characteristics. Columns (3) to (6) feature a significantly lower number of observations because of limited data coverage for labor union, entrenchment index, and BoardEx variables. Detailed variable definitions are in Appendix B. Industry classifications are based on Fama-French 48 industries. Standard errors are clustered at the firm level.

	(1)	(2)	(3)	(4)	(5)	(6)
Size (log assets)				.097***	.097***	.112***
				(.009)	(.009)	(.024)
ROA				.044	.044	.029
				(.159)	(.159)	(.134)
Sales growth				.128**	.128**	.072
				(.06)	(.06)	(.053)
Tobin's q				.057***	.057***	.034***
				(.01)	(.01)	(.008)
Total debt/assets				-.338***	-.338***	-.08
				(.064)	(.064)	(.084)
cash/assets				.377***	.377***	-.011
				(.093)	(.093)	(.106)
Labor union coverage				-.049	-.049	-.41
				(.308)	(.308)	(.311)
Entrenchment index				.013	.013	-.019
				(.015)	(.015)	(.017)
Board nationality mix						.12
						(.082)
Board gender ratio						-.63***
						(.098)
Board age diversity						-.009
						(.006)
Board diversity: firm tenure						0
						(.005)
Board: network size						0**
						(0)
Board diversity: qualifications						.022
						(.027)
_cons	3.436***	3.437***	3.438***	2.494***	2.494***	3.18***
	(.009)	(.009)	(0)	(.109)	(.109)	(.246)
Observations	19996	19962	24657	7884	7884	7156
R-squared	.035	.099	.38	.214	.214	.437
Fixed effects	Industry	Industry*Year	Firm	Industry*Year	Industry*Year	Firm

Standard errors are in parentheses
*** p<.01, ** p<.05, * p<.1*