

Are Firms Sacrificing Flexibility for Diversity and Inclusion?

Hoa Briscoe-Tran*

March 13, 2026

Abstract

This paper examines how workforce diversity and inclusion (D&I) affect corporate flexibility – a firm's ability to adapt to changes, especially 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 some evidence suggesting a causal interpretation. 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: Corporate 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.

* Monash School of Business, Monash University (email: hoa.briscoe-tran@monash.edu). This paper has benefited from comments by Alex Edmans, Jocelyn Evans, Isil Erel, Laura Field, Simon Glossner (discussant), John Graham, Isaac Hacamo, Davidson Heath, Susan Jackson, Mahsa Kaviani (discussant), 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, the 2024 University of Connecticut Finance Conference, 2025 American Economic Association (AEA) Annual Meeting, and the 2025 ECGI Corporate Governance Symposium and University of Delaware John L. Weinberg/IRRCi Research Paper Award. 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 corporations. This shift extends beyond the top management or board level to the broader workforce level, 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 a firm's operations. This issue has become highly debated, particularly amid the recent political backlash against D&I initiatives across the United States, raising the question of their economic trade-offs.

A review of economic theories highlights several potential costs of D&I practices. Diversity may create communication challenges (Lang (1986)) and increase social division and miscoordination among workers (e.g., Alesina, Baqir, and Easterly (1999), Mannix and Neale (2005)). Furthermore, fostering inclusiveness among diverse agents often takes time (e.g., Jackson (1992)), which could slow decision-making and implementation. In addition, D&I practices that target hiring, firing, and promotion decisions may constrain a firm's workforce management (Athey, Avery, and Zemsky (2000), Fryer Jr and Loury (2013)), effectively increasing labor adjustment costs. Taken together, these frictions may reduce corporate flexibility—a firm's ability to adapt effectively to economic changes.

Nonetheless, there are reasons to expect the opposite. A diverse workforce could provide an organization with more information and approaches to problem-solving (Ibarra (1993), Ibarra (1995), Ibarra (1997)), enabling better decisions in adapting to changes. Experimental evidence also suggests that diverse groups process information more effectively and exhibit more critical thinking (Hoffman (1959), Hoffman and Maier (1961), Nemeth (1986)). Moreover, D&I practices can help a firm gather more customer support (Hacamo (2023)) and attract more prospective employees (Avery and McKay (2006), Choi et al. (2023)), allowing not only more flexible expansion in good times but also greater adaptability in downturns, as loyal customers ease contractions and deeper labor pools make future rehiring less costly. Overall, whether D&I strengthens or weakens corporate flexibility is an empirical question.

The focus on corporate flexibility arises naturally from theories about D&I's economic trade-offs and addresses a vital aspect of firm success. Barry, Campello, Graham, and Ma (2022) highlight corporate flexibility's importance for firms' survival during crises, while many studies

show its significant impact on firm risk, returns, and financial policies.¹ The global pandemic has heightened awareness of corporate flexibility's importance, yet its drivers remain underexplored.

Research on diversity and inclusion (D&I) at the firm level, however, faces substantial data limitations. Among many issues, Edmans, Flammer, and Glossner (2023) highlight that the existing literature centers on demographic diversity while giving limited attention to inclusion. Using confidential survey data for a few hundred firms, they show that inclusion correlates little with diversity metrics, underscoring the need for measures that capture both aspects. Building on this insight, we construct a new measure of D&I from public and larger data – over 10 million employee reviews across thousands of public and private companies – thereby opening new possibilities for future research. Specifically, we draw on a pioneering D&I rating introduced in 2020 by a leading career intelligence platform, designed to reflect employees' perspectives on workplace diversity and inclusion. Leveraging this rating, we extrapolate it back to 2008 using review texts and a large language model from Google. Validation tests confirm that our measure captures D&I-related constructs as intended: it places positive weight on inclusive language and negative weight on discriminatory language, and it strongly predicts whether a firm is recognized on *Fortune's Best Companies for Diversity* lists as voted by women, LGBTQ employees, people of color, and workers with disabilities.

Using our D&I measure, we examine its relationship with corporate flexibility in a sample of over 2,000 large U.S. publicly listed firms. Following Gu, Hackbarth, and Johnson (2018), we measure flexibility as the historical range of operating cost margins scaled by the volatility of productivity growth. The idea is that flexible firms adjust operations more readily in response to productivity shocks—expanding when conditions improve and contracting when they worsen. Under diminishing returns to scale, expanding after positive shocks or contracting after negative ones partially offsets the effect of those shocks on cost margins, keeping them more stable. In contrast, inflexible firms delay adjustment, allowing margins to swing more widely. Thus, a wider range indicates lower flexibility due to weaker shock absorption.

Our analysis shows that firms with higher D&I ratings exhibit significantly lower flexibility during 2009–2021: one standard deviation (SD) higher D&I rating is associated with 8.7% SD

¹ Risk and return (Hackbarth and Johnson (2015), Gu, Hackbarth, and Johnson (2018), Fisher, Knesl, and Lee (2025)), 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)).

lower flexibility. This effect remains significant after controlling for known determinants of flexibility, including fixed assets ratio, production technology (proxied by industry-by-year fixed effects), labor regulations and labor market conditions (proxied by state-by-year fixed effects), and other factors such as governance and top management and board diversity metrics. The results remain statistically significant after correcting for estimation uncertainty due to the fact that our D&I rating is estimated. Moreover, the results are similar for expansion flexibility and contraction flexibility, suggesting that the lower flexibility of D&I firms reflects overall adjustment inertia rather than asymmetric responses to positive or negative shocks.

While the negative association between D&I ratings and corporate flexibility appears robust, establishing a causal interpretation is inherently challenging. D&I practices are endogenous, and both D&I and flexibility are persistent traits that may reflect unobserved firm characteristics. To mitigate these concerns, we first control for flexibility measured in the period preceding our main sample, helping to account for persistent unobservable firm-level attributes that shape corporate flexibility. The results remain unchanged. Second, we consider plausible theories on the endogenous nature of D&I practices. The most relevant framework is by Ferreira and Nikolowa (2025), who model nonpecuniary workplace attributes – such as purpose, sustainability, and D&I – as strategic tools firms use to attract workers with heterogeneous social preferences. In their model, firms with more flexible production technologies can better tailor job attributes to match extreme worker preferences, thereby maximizing surplus. Consequently, flexible firms are more likely to have either very high or very low D&I ratings, depending on which end of the preference spectrum they target. This would predict a U-shaped relationship between flexibility and D&I. However, our data reveals a monotonic negative relationship, inconsistent with this prediction.

Finally, despite the persistence and endogenous nature of D&I, our sample includes a plausibly exogenous shock to firms' D&I practices originating from a Supreme Court ruling. Under Title VII of the Civil Rights Act in the U.S., employer liability for workplace harassment depends on whether the harasser is classified as a supervisor. Prior to 2013, U.S. circuit courts applied different definitions, with some adopting a broader standard that increased firms' exposure to harassment liability. In *Vance v. Ball State* (2013), the Supreme Court resolved this split by adopting a narrow definition of supervisors, forcing circuits with the broader standard to reduce employer liability. The decision was closely divided and difficult to anticipate, and legal studies document a substantial decline in successful harassment claims following the ruling. By lowering firms' legal

exposure to harassment, the decision likely weakened incentives to maintain (sexual and racial) harassment prevention and, by extension, related inclusive workplace practices. We therefore examine whether firms exposed to the ruling subsequently experience declines in D&I ratings relative to other firms and whether corporate flexibility evolves in tandem with these changes.

Our difference-in-differences analysis reveals that exposed firms experienced a significant decline in their D&I ratings and an increase in flexibility relative to control firms following the ruling. The ruling, however, may have triggered multiple organizational changes, ranging from reductions in legal departments to generally less cautious decision-making in a less litigious environment. While these concurrent effects make it challenging to isolate the pure impact of D&I on flexibility, the analysis remains valuable for two reasons. First, it helps rule out many alternative explanations related to unobservable firm characteristics or managerial ability. Second, even if some of the flexibility increase stems from broader organizational responses to decreased litigation risk, the results indicate that D&I considerations, at least through legal requirements, affect corporate flexibility.

Two potential channels may explain why D&I negatively impact corporate flexibility. The *workforce channel* suggests that D&I considerations create hiring and firing frictions that restrict workforce management. The *decision-making channel* posits that diverse groups face communication challenges and slower consensus formation, hindering agile adaptation to changes.

To test these channels, we examine how D&I firms adjusted their workforce and firm-level decision-making responses during the 2020 COVID crisis – an ideal shock for observing adjustment flexibility. Using firms' reliance on in-person interactions as a measure of COVID exposure (Koren and Peto (2020)), we find D&I firms' operating performance was more sensitive to COVID exposure, indicating lower adaptability. This stems from efficiency rather than workforce management: while D&I firms' workforce adjustments matched other firms, they saw larger declines in operating efficiency (sales per employee, sales growth, and assets turnover). These findings persist after controlling for financial conditions, governance, board diversity, customer base, labor characteristics, and industry trends. The results suggest D&I primarily affects flexibility through operational efficiency frictions rather than workforce constraints.

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 (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 more well-defined issue that has recently attracted substantial media attention. Importantly, 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 growing D&I focus, firm-level measurement has been limited. While Edmans, Flammer, and Glossner (2023) created a novel measure using D&I-relevant questions from Fortune's Best Companies surveys, our approach differs in three key ways: (1) we use direct employee perceptions of D&I practices, which align with general D&I definitions, rather than researchers' judgment of relevant survey items, (2) we analyze over 2000 firms versus their smaller sample (limited by Fortune competition participation), and (3) we examine D&I's causal effect on firm flexibility rather than accounting and stock performance.

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 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 labeling is time-consuming and generates limited training data. Our approach differs by having BERT learn directly from employees' provided labels, 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 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

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

(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. We 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, we 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.

While the majority of reviews have all the typical information listed above, only the overall rating, the pros and cons sections, and the review title are mandatory for every review. The other ratings are not mandatory: 79.7% of the reviews have all of them available (see the Internet Appendix Table IA1 for the review-level summary statistics).

³ See <https://www.glassdoor.com/blog/diversity-inclusion-products/>.

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 the model to understand review texts, we use BERT (Bidirectional Encoder Representations from Transformers), which have been shown to outperform 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 each sentence as a whole. More importantly, BERT is pre-trained using a large corpus (3.3-billion-word corpus from English books and Wikipedia), so it has gained significant knowledge of the English language. Consequently, by fine-tuning the BERT model, i.e., adjusting the model’s weights to fit better to our setting of predicting the D&I rating based on review texts, we can transfer the model’s general knowledge of the English language to the specific context of employee reviews.

To apply the BERT model in our setting, we proceed in three steps. First, we combine the relevant information in each review into a single text that BERT can process. Appendix A shows how an example review is represented as one text containing its pros, cons, title, and all ratings.

Second, we feed the combined text into a simplified version of BERT, called distilled BERT (Sanh et al. (2019)), to reduce computational burdens. In doing so, we instruct the model to predict the D&I rating in a classification task with five classes representing five possible ratings: one to five stars. We train and test 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, because not all reviews include every numerical rating beyond the overall rating, we train two main models: a simpler model that uses only the pros, cons, title, and overall rating, which are available for all reviews, and a fuller model that incorporates all available details, including additional numerical ratings, but is restricted to the 79.7% of reviews where these are present. The out-of-sample D&I rating for each review is then taken from the fuller model whenever the data allows, and from the simpler model otherwise.

Regarding performance, the simpler model achieves 54.55% accuracy, while the fuller model achieves 63.83%, compared to 20% for random guessing in a five-class classification task. This level of model performance is in line with the performance of other five-class classifications in the machine learning literature (e.g., between 59% and 70% accuracy for predicting 5-star Yelp ratings as in Liu (2020), though comparisons should be made with caution. Our task predicts the D&I rating from Glassdoor reviews—a subcategory less directly aligned with review text than overall ratings—making it inherently more challenging. Also, we use distilled BERT due to computational constraints, while prior studies often rely on larger models like BERT base or RoBERTa.

For completeness, we subsequently estimate a third model that applies the fuller specification to the entire sample using whatever information is available. While this raises accuracy to 64.26% (a modest 0.43% gain), it risks inconsistency by treating reviews with missing information the same way as those with non-missing information. So, we retain this third model only as a robustness. For this iteration of model training, we also save the full probability distribution across classes to capture BERT prediction uncertainty and how that may affect our later regressions.

One concern with our BERT application is that it does not explicitly model the ordinal nature of the ratings (e.g., that a 1-star rating is closer to a 2-star rating than to a 5-star rating). The machine learning literature suggests potential gains from incorporating ordinality (Gutierrez et al. (2016)), though evidence is mixed (Ben-David, Sterling, and Tran (2009)). In response, we train an ordinal variant of BERT following Sheshadri, Vijjini, and Kharbanda (2021), using the same input structure as the third model above so that comparisons are made directly to that benchmark. The ordinal variant augments the standard classification loss with a distance-penalizing term, giving larger misclassifications heavier penalties. We evaluate performance using both accuracy, which is standard in classification tasks, and correlation with the true labels, which also reflects ordinal proximity. On a hold-out test set, the ordinal variant records an improvement in correlation by 0.0008 from 0.7713, but also a small drop in accuracy by 0.02%; the two sets of predictions correlate at 0.9826. Given these nearly identical results, we retain the baseline for parsimony and replicability, with full details of the ordinal model in the Internet Appendix (IA2).

We use the trained models to impute the D&I rating for reviews without an actual D&I rating while retaining the reported rating whenever available (only after 2020). Each firm’s annual D&I rating is then the average of imputed and actual ratings across all reviews in each year.

3.3. Validating the measure

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

3.3.1. At the review level

Our D&I rating ultimately relies on employees' judgments of firms' D&I practices. However, if employees misunderstand D&I or conflate it with other factors (e.g., job satisfaction), the rating may not accurately capture D&I. To address this concern, we implement two complementary checks: (1) an interpretability check of what the rating reflects, and (2) a validation check against an independent text-based proxy.

For the interpretability check, we adopt the approach of Acikalin et al. (2022) to identify the most informative words about our 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.

For the second validation, we adopt the bag-of-words methodology in Hanley and Hoberg (2019), Li et al. (2021), Briscoe-Tran (2022), to construct a D&I-specific measure for each review based on the relative frequency of D&I keywords in the pros versus cons sections. Although this word-count approach is a simplified proxy, it provides a benchmark for whether our measure aligns with explicit D&I language. The correlation between this measure and our D&I rating is 0.25, compared to 0.21 for human ratings on a hold-out testing sample, with similar results across model variants. This suggests that our models capture D&I-related textual signals, providing reassurance that the rating aligns with explicit D&I language.

3.3.2. At the firm-year level

We take the average D&I rating across reviews in each firm-year to be our main D&I measure at the firm-year level.⁴

Ideally, to validate this measure, we 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

⁴ We 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, we find in un-tabulated tests that this relationship is not robust after controlling for our main D&I measure.

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), 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 our D&I measure, akin to an out-of-sample test.

Table 2 Panel A shows that our 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 too, [tabulated in Internet Appendix Table IA5](#)). 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 our D&I measure remains statistically significant after controlling for the Refinitiv rating (column (6)). The economic magnitude of the coefficient on our 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, our 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 our 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 is 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,

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 our D&I rating would be $\exp((3.855-3.036)*2.852) = 10.33$.

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, we 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, we collect the top 100 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, we investigate whether our 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 our rating. We find that to be the case, as shown in Table 2 Panel C. Our 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, our 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, our D&I measure better captures the firm's grassroots D&I practices.

Finally, we evaluate the persistence in our 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. We 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 we restrict our sample to firm-years with above-median number of reviews, where idiosyncratic noise is likely less (column

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/>

(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 our D&I measure shows strong persistence within each firm, but its persistence also captures unique information beyond other ratings on Glassdoor.

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

3.4. Summary statistics

While we have reviews covering over 300,000 employers listed on Glassdoor, in this paper we focus on a sample of the largest publicly listed companies in the US because these firms report abundant data on operating performance. In particular, we restrict our 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 we construct the sample and match that to other databases like Compustat for accounting data.

We 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 our sample).

In light of these correlations, we 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 our 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 measure of corporate flexibility from the literature. While flexibility has proven difficult to quantify (Chen, Kacperczyk, and Ortiz-Molina (2011)), Gu, Hackbarth, and Johnson (2018) propose a theoretically grounded metric based on the historical range of a firm's operating cost margins relative to fluctuations in productivity—measured by changes in sales over assets.

The core idea is that when productivity shocks occur, inflexible firms are slower to adjust operations. This delay allows operating cost margins to reach more extreme values, resulting in a wider range. For instance, following a positive shock, a firm's profit margins rise. A flexible firm would respond by increasing production to capitalize on the shock. As production scales up, diminishing returns temper the profit margin increase, keeping profit and cost margins relatively stable. In contrast, an inflexible firm may not adjust production, allowing profit margins to remain elevated. The reverse holds for negative shocks. Consequently, inflexible firms tend to exhibit a broader range of profit and cost margins in response to productivity changes.

Using that measure of flexibility, we first provide cross-sectional evidence on the relationship between a firm's D&I rating and its flexibility. After that, we provide evidence on how a firm's flexibility changes following a plausibly exogenous shock to D&I.

4.1. Cross-sectional evidence

Following Gu, Hackbarth, and Johnson (2018), estimating corporate flexibility requires long accounting histories, as estimation quality increases with the number of years of data. Because our D&I ratings are available between 2009 and 2021, we use accounting data from the same 2009–

2021 period to compute each firm's flexibility.⁹ This design naturally lends itself to a cross-sectional test rather than a time-series analysis, since both flexibility and D&I are persistent firm characteristics, and the multi-year estimation of flexibility further reinforces that persistence.^{10 11}

Although flexibility is estimated once per firm across the full period, D&I ratings vary annually. To exploit this panel structure, we assign each firm's flexibility estimate to every year in the panel and regress it on the annual D&I rating and other firm-year characteristics. This setup maximizes the precision of the flexibility measure while allowing us to account for temporal variations in D&I. Because the dependent variable is repeated for each firm, we cluster standard errors at the firm level to account for serial correlation.

Formally, we estimate:

$$FLEX_i = \beta_0 + \beta_1 * D\&I_{i,t} + B_2 * X_{i,t} + v_{j,t} + u_{s,t} + \epsilon_i$$

Where $FLEX_i$ is the inverse of inflexibility, defined as the range of operating cost margins over the standard deviation of log changes in sales over assets, calculated over 2009–2021 and repeated for each year. $D\&I_{i,t}$ is firm i 's D&I rating based on Glassdoor reviews during year t , up to 2021, when we downloaded Glassdoor data. $X_{i,t}$ denote various control variables, while $v_{j,t}$ and $u_{s,t}$ denote industry-year and state-year fixed effects, respectively; ϵ_i is the error term.

Control variables include factors that may influence corporate flexibility. First, a firm's production technology could shape its flexibility, so we include industry-year fixed effects and other proxies for production function characteristics, including fixed asset ratio (net property, plant, and equipment scaled by total assets) and labor stock (number of employees relative to total assets). Second, we add state-year fixed effects to absorb variation in local labor regulations and workforce characteristics, since employment protection laws and workforce composition or preferences may also constrain corporate flexibility. Third, we account for firm-specific attributes that may be correlated with D&I and could also affect responsiveness to shocks, including firm size (log of assets), corporate governance attributes (institutional and insider ownership), and

⁹ We drop 2008, the inception year of Glassdoor, because reviews were sparse and only available for part of that year.

¹⁰ 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 Table IA6, we 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.

¹¹ We 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. Our results remain similar when FLEX is measured with quarterly data (Table IA6).

board and top management diversity metrics. We do not control for cash holdings or capital structure in our baseline, as these are typically downstream outcomes of flexibility rather than determinants.¹² To reduce noise, all ratio variables are winsorized at the 1st and 99th percentiles, and firms with fewer than five years of cost margin data are excluded.

The results in Table 3 indicate that D&I firms exhibit significantly lower flexibility than other firms. Across the columns, the coefficient on the D&I rating indicates that a one-standard-deviation increase in D&I is associated with a 5.5% to 8.7% standard-deviation decrease in flexibility. This effect is statistically significant at the 1% level and remains so after sequentially adding controls for labor stock, fixed asset ratio, firm size (column 2), industry-year and state-year fixed effects (columns 3 and 4), and governance variables such as institutional ownership, institutional ownership concentration, and insider ownership (column 5).

These findings remain robust across multiple specifications, as shown in Internet Appendix Table IA6. Specifically, the results hold when excluding firms with fewer than ten years of cost margin data and when restricting the sample to firm-years with at least five or ten Glassdoor reviews. They also persist after controlling for governance-related factors, including the Bebchuk, Cohen, and Ferrell (2009) entrenchment index – omitted from the main tests because it is available only for a subset of firms – and corporate D&I policies from Refinitiv, which are excluded from the main tables as they are only available post-2016. Importantly, the coefficient on D&I remains strong after incorporating board and top management diversity measures (gender, nationality, tenure, age, and network size from BoardEx), indicating that workforce D&I effects are distinct from leadership diversity.

Finally, because the D&I rating is estimated with uncertainty, our baseline standard errors may understate true variability. To address this, we implement two robustness checks. First, we compute firm-level D&I ratings using probability-weighted averages from a full distribution of BERT model outputs rather than relying solely on the most likely label. Second, we apply a bootstrap procedure that simulates D&I ratings for each review based on the BERT-predicted label probabilities, aggregates these into firm-year averages, and resamples both reviews within firm-years and firm-years themselves. This approach captures uncertainty from the BERT estimation,

¹² See, e.g., Ghaly, Anh Dang, and Stathopoulos (2017), Simintzi, Vig, and Volpin (2015), Serfling (2016), Reinartz and Schmid (2016), D’Acunto et al. (2018), Gu and Hackbarth (2021). Results are qualitatively the same with cash and debt as controls (shown in Internet Appendix Table IA6).

the averaging across reviews, and the regression estimation altogether. As shown in Internet Appendix Table IA7, these adjustments do not materially affect the results. Given the computational demands of this three-layer simulation, we do not use it as the baseline and keep details in the Internet Appendix.

4.2. Alternative explanations and a plausibly exogenous shock

While the negative association between D&I and corporate flexibility is robust, interpreting this relationship as causal is challenging. D&I practices are not randomly assigned, and both D&I and flexibility are persistent firm attributes that may reflect unobserved characteristics. In this section, we conduct a series of analyses designed to strengthen the causal interpretation.

First, although firm fixed effects are not feasible in our baseline – since flexibility is inherently persistent and operationally estimated only once per firm – we can still account for time-invariant firm characteristics relevant for flexibility by controlling for historical flexibility estimated over a sufficiently long pre-sample period. Controlling for this pre-period flexibility in the regression does not alter the results, as reported in Internet Appendix Table IA8.

Second, we examine plausible reverse-causality channels in which corporate flexibility may shape D&I rather than the other way around. One such hypothesis draws on the idea that D&I, as part of corporate social responsibility, may be viewed as a luxury good (e.g., Masulis and Reza (2015); Bansal, Wu, and Yaron (2022)). Under this view, flexible firms may scale back D&I efforts during downturns when resources are tight. If our sample contains proportionally more contraction periods than expansion periods, this mechanism could generate the observed negative association between flexibility and D&I.

To evaluate this possibility, we compute flexibility separately in expansion vs. contraction periods, defined by whether annual productivity growth (changes in sales over assets following Gu et al. 2018’s methodology) is positive or negative. If the D&I-as-a-luxury explanation were driving the results, we would expect the negative association to arise primarily in contraction periods. However, as shown in Table 4, the relationship between D&I and flexibility remains negative and statistically significant when flexibility is calculated within expansion years and within contraction years separately, with economic magnitudes even slightly larger in expansion years. This pattern indicates that D&I firms exhibit lower flexibility regardless of the direction of productivity shocks, inconsistent with the luxury-good explanation.

Another theory of endogenous D&I practices is provided by Ferreira and Nikolowa (2025), who model nonpecuniary job attributes – such as purpose, sustainability, and D&I – as tools firms use to attract workers with heterogeneous social preferences. In their framework, production technology interacts with worker preferences to shape firms’ optimal D&I choices. Firms with flexible production technologies can more easily tailor job attributes to appeal to workers at the extremes of the preference distribution, thereby extracting more surplus from these workers. This would imply a U-shaped prediction: highly flexible firms should be more likely to adopt either very high or very low levels of D&I. In contrast, inflexible firms gravitate toward moderate levels.

Our data instead shows a monotonic negative relationship between flexibility and D&I (Figure 1), inconsistent with the U-shaped prediction. Still, the Ferreira–Nikolowa framework underscores potential determinants of D&I adoption that we incorporate in our design: industry-by-year fixed effects to capture production technology heterogeneity and state-by-year fixed effects to proxy for regional variation in worker preferences. These baseline controls help ensure that the negative relationship is not driven by differences in technology or workforce composition across regions.

Finally, we exploit a plausibly exogenous shock to firms’ D&I practices arising from an unexpected change in U.S. employment law affecting workplace harassment prevention. Such a shock could affect firms’ D&I practices because harassment prevention is part of firms’ broader employee-protection and inclusion efforts, and harassment cases frequently involve racial and gender issues that are central to D&I assessments.

Under Title VII of the Civil Rights Act of 1964, employers can be held vicariously liable for workplace harassment committed by employees with supervisory authority. Prior to 2013, U.S. circuit courts differed in how they defined a “supervisor.” The First, Seventh, and Eighth Circuits applied a narrow standard – only employees with the power to hire, fire, promote, demote, transfer, or discipline qualified. Under this view, harassment by low-level supervisors or those directing daily work triggered employer liability only if the employer was negligent. In contrast, the Second, Fourth, and Ninth Circuits used a broader definition, classifying anyone who oversaw another’s daily work as a supervisor, making it easier to hold employers vicariously liable.

The Supreme Court resolved this circuit split in its 2013 decision in *Vance v. Ball State University*, which adopted the narrow definition of supervisory status. As a result, firms in circuits that had previously applied the broader standard experienced a sudden reduction in liability exposure and weaker legal incentives to prevent and remediate workplace harassment. The ruling

was decided by a narrow 5-to-4 margin, limiting scope for anticipation. A scholarly legal analysis by Sheldon-Sherman (2021) documents that the decision led to an approximately 20% decline in successful harassment claims, highlighting the significance of the *Vance* ruling.

To the extent that firms' investments in harassment prevention are linked to broader employee-protection and inclusion practices, this legal change could be associated with a decline in D&I ratings for affected firms (firms headquartered in the Second, Fourth, and Ninth Circuits) relative to other firms (firms headquartered in the First, Seventh, and Eighth Circuits) following the ruling.¹³ We then examine whether such changes in D&I are accompanied by corresponding changes in corporate flexibility. Our empirical analysis follows a standard difference-in-differences design with two-way fixed effects. Given the objective of capturing short-term adjustment dynamics around the shock, we re-estimate corporate flexibility using rolling three-year windows. Longer estimation windows would mechanically smooth short-run variations. Following recent guidance on difference-in-differences designs, we do not include time-varying firm controls, which may themselves be affected by the treatment and thus introduce post-treatment bias (see Abadie et al. (2025) and references therein). Standard errors are clustered at both the state level (because our treatment is at the state level) and the firm level (to account for within-firm overlapping estimation window for the flexibility metric).

Table 5 reports the results. Relative to the year before the court ruling, treated firms experienced, on average, a 0.104 standard deviation decline in D&I ratings in the five years following the ruling compared to control firms, statistically significant at the 10% level (column 1). Column 2 shows that the estimated decline was strongest in the first year of exposure to the ruling (a 0.234 standard deviation effect, significant at the 1% level). Column 3 shows that treated firms also experienced, on average, a relative increase in short-term corporate flexibility in the five years following the ruling compared to the pre-period, amounting to a 0.155 standard deviation increase and statistically significant at the 1% level. Column 4 reports dynamic estimates, with the largest coefficient corresponding to three years after the shock, consistent with an increase in flexibility measured over the three-year window ending three years after the ruling.

¹³ Other circuits lacked clearly established precedents on the definition of supervisory status prior to 2013 (Sheldon-Sherman (2021)) and are therefore excluded from the analysis. We assign exposure based on firms' headquarters location, which proxies for firms' primary employment jurisdiction and centralized policy responses to legal risk.

Figure 2 provides graphical evidence consistent with the above analysis and allows for an assessment of pre-treatment dynamics. For D&I ratings, there is no evidence of significant differential pre-trends between treated and control firms in the three years prior to the court ruling. For corporate flexibility, there is some evidence of a pre-treatment difference, though it is opposite in sign to the post-ruling effect and therefore does not suggest that treated firms were on an increasing flexibility trend prior to the shock. Nonetheless, to ensure that these pre-treatment differences do not affect the results, we implement a nearest-neighbor matching procedure in which each treated firm is matched to the closest control firm based on changes in flexibility in the two and three years prior to the ruling, relative to the year immediately before the ruling. In the matched sample, pre-treatment differences are no longer present, while the main results remain similar, as shown in Internet Appendix Figure IA3.

The court ruling likely induced organizational responses beyond changes in D&I practices, including adjustments to firms' legal compliance and risk management functions, which makes it difficult to isolate a single mechanism linking the shock to changes in corporate flexibility. Nevertheless, the analysis remains informative. The ruling represents an external legal shock arising from litigation outside the corporate sector and unrelated to firms' own actions, which helps rule out a broad set of alternative explanations based on endogenous firm decisions. Moreover, even if most of the observed response reflects broader organizational adjustments to reduced litigation risk, the results indicate that legal incentives related to workplace inclusion and employee protection are relevant for shaping corporate flexibility.

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 our 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. We measure operating performance by return on assets (ROA). We 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, we 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, we 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, we 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, we conduct similar tests to the previous section but instead focus on workforce cuts and measures of efficiency. For efficiency, we 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 we 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, we 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 our 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, our 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 we cannot control for labor union directly due to the lack of firm-level union data, we 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, we 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

- Abadie, Alberto, Joshua Angrist, Brigham Frandsen, and Jörn-Steffen Pischke, 2025, Harvesting Differences-in-Differences and Event-Study Evidence, *National Bureau of Economic Research*.
- 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.
- 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*.
- Bansal, Ravi, Di Wu, and Amir Yaron, 2022, Socially Responsible Investing in Good and Bad Times, *The Review of Financial Studies* 35, 2067–2099.
- 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).
- Ben-David, Arie, Leon Sterling, and TriDat Tran, 2009, Adding Monotonicity to Learning Algorithms May Impair Their Accuracy, *Expert Systems with Applications* 36, 6627–6634.
- 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.

Ferreira, Daniel, and Radoslaw Nikolowa, 2025, Polarization, purpose and profit, *Journal of Financial Economics* 172, 104147.

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.

Fisher, Adlai, Jiří Knesl, and Ryan CY Lee, 2025, How Valuable Is Corporate Adaptation to Crisis? Estimates from Covid-19 Work-from-Home Announcements, *Journal of Financial Economics* 174, 104168.

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, .

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.

Gutierrez, Pedro Antonio, Maria Perez-Ortiz, Javier Sanchez-Monedero, Francisco Fernandez-Navarro, and Cesar Hervás-Martinez, 2016, Ordinal Regression Methods: Survey and Experimental Study, *IEEE Transactions on Knowledge and Data Engineering* 28, 127–146.

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.

Hanley, Kathleen Weiss, and Gerard Hoberg, 2019, Dynamic Interpretation of Emerging Risks in the Financial Sector, *The Review of Financial Studies*.

- 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.
- 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.
- Li, Kai, Feng Mai, Rui Shen, and Xinyan Yan, 2021, Measuring Corporate Culture Using Machine Learning, *The Review of Financial Studies* 34, 3265–3315.
- Liu, Zefang, 2020, Yelp Review Rating Prediction: Machine Learning and Deep Learning Models, (arXiv).

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, 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.

Masulis, Ronald W., and Syed Walid Reza, 2015, Agency Problems of Corporate Philanthropy, *Review of Financial Studies* 28, 592–636.

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.

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.

Sheldon-Sherman, Jennifer AL, 2021, The Effect of Vance v. Ball State in Title VII Litigation, *U. Ill. L. Rev.*, 983.

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

Sheshadri, Akshay Krishna, Anvesh Rao Vijjini, and Sukhdeep Kharbanda, 2021, WER-BERT: Automatic WER Estimation with BERT in a Balanced Ordinal Classification Paradigm, (arXiv).

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 our 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 assets)	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/assets	-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 executives	0.139***	1.000				
(3) Refinitiv board	0.186***	0.359***	1.000			
(4) Refinitiv board diversity	0.060***	0.073***	0.113***	1.000		
(5) Refinitiv D&I score	-0.077***	-0.070***	-0.073***	0.022	1.000	
(6) Refinitiv D&I score	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 - compensation	0.556***	0.679***	0.522***	0.657***	0.654***	1.000	
(7) Rating - management	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 our 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*** (.176)	2.338*** (.246)	1.987*** (.233)	3.003*** (.352)	2.423*** (.368)	2.852*** (.43)	2.737*** (.624)	1.637** (.646)
Lagged Best Diversity			5.284*** (.355)		4.776*** (.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*** (.203)	2.066*** (.338)	1.863*** (.334)	2.228*** (.5)	2.001*** (.504)	2.14*** (.526)	2.801*** (.686)	2.818*** (.772)
Fortune's Best Diversity ₂₀₁₉			1.756*** (.379)		1.011** (.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*** (.085)	.838*** (.104)	.442*** (.131)	.695*** (.213)	.157 (.229)	.591 (.468)	.663*** (.176)	1.581*** (.225)
Lagged DiversityInc			7.445*** (.275)		6.357*** (.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*** (.01)	.404*** (.011)	.04*** (.005)	.714*** (.008)	.691*** (.01)	.068*** (.005)
Rating - overall			.397*** (.028)			.422*** (.014)
Rating - balance			.059*** (.011)			.014* (.008)
Rating - culture			.369*** (.016)			.462*** (.011)
Rating - career			.037*** (.013)			.027** (.011)
Rating - compensation			-.033*** (.008)			-.072*** (.007)
Rating - management			.093*** (.014)			.01 (.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) on annual D&I ratings and other firm-year characteristics. FLEX is estimated once per firm over 2009–2021 as the ratio of the standard deviation of log sales-to-assets growth to the range of operating cost margins (Gu et al., 2018), and assigned to all years for that firm in the panel. The regressors include the D&I rating, labor stock, fixed assets ratio, and size, with additional specifications adding ownership and governance variables. All continuous variables are standardized to have a zero mean and a unit standard deviation. All ratio variables are winsorized at the 1st and 99th percentiles, and firms with fewer than five years of cost margin data are excluded. Industry classifications follow Fama-French 48 industries. State denotes a firm's headquarters state. Standard errors are clustered by firm.

	(1) FLEX	(2) FLEX	(3) FLEX	(4) FLEX	(5) FLEX
D&I rating	-0.087*** (0.013)	-0.079*** (0.014)	-0.064*** (0.014)	-0.061*** (0.014)	-0.055*** (0.014)
Labor stock		0.074*** (0.021)	0.050*** (0.018)	0.041** (0.018)	0.031* (0.018)
Fixed assets ratio		-0.039** (0.020)	-0.146*** (0.042)	-0.144*** (0.044)	-0.127*** (0.037)
Size		0.012 (0.025)	0.095*** (0.029)	0.075** (0.030)	0.070** (0.031)
Inst. ownership					-0.011 (0.023)
Inst. ownership HHI					-0.058** (0.024)
Inside ownership					-0.019 (0.013)
_cons	-0.003 (0.023)	-0.028 (0.023)	-0.028 (0.021)	-0.025 (0.021)	-0.045*** (0.021)
Observations	20879	16872	16837	16333	13540
R-squared	0.008	0.014	0.165	0.215	0.218
Industry*year FE	No	No	Yes	Yes	Yes
State*year FE	No	No	No	Yes	Yes

Standard errors are in parentheses

*** $p < .01$, ** $p < .05$, * $p < .1$

Table 4: Flexibility and D&I in expansion and contraction periods

This table tests whether the association between Diversity & Inclusion (D&I) and corporate flexibility (FLEX) depends on the economic environment. FLEX is re-estimated separately for years in which a firm experiences positive versus negative annual productivity growth, following Gu et al. (2018). For each firm, we construct two measures: flexibility computed only from contraction years and flexibility computed only from expansion years. Each measure is then merged into the panel and analyzed using only the corresponding subset of firm-year observations (i.e., regressions for contraction FLEX are estimated on contraction-year observations, and regressions for expansion FLEX are estimated on expansion-year observations). Apart from conditioning on the relevant state (expansion or contraction), all specifications mirror those in Table 3.

Panel A: Restricting all measures to expansion years

	(1) Expansion FLEX	(2) Expansion FLEX	(3) Expansion FLEX	(4) Expansion FLEX	(5) Expansion FLEX
D&I rating	-0.094*** (0.020)	-0.087*** (0.017)	-0.073*** (0.017)	-0.070*** (0.020)	-0.052** (0.021)
Labor stock		0.067** (0.028)	0.050* (0.027)	0.043 (0.026)	0.045* (0.027)
Fixed assets ratio		-0.046** (0.019)	-0.128*** (0.030)	-0.135*** (0.032)	-0.141*** (0.034)
Size		0.061*** (0.021)	0.092*** (0.024)	0.076*** (0.026)	0.085*** (0.028)
Inst. ownership					-0.002 (0.024)
Inst. ownership HHI					0.010 (0.047)
Inside ownership					-0.028* (0.015)
_cons	0.005 (0.026)	-0.038* (0.021)	-0.037* (0.020)	-0.031 (0.020)	-0.041* (0.022)
Observations	11216	7689	7620	7325	5949
R-squared	0.009	0.016	0.136	0.205	0.207
Industry*year FE	No	No	Yes	Yes	Yes
State*year FE	No	No	No	Yes	Yes

Standard errors are in parentheses

*** $p < .01$, ** $p < .05$, * $p < .1$

Panel B: Restricting all measures to contraction years

	(1) Contraction FLEX	(2) Contraction FLEX	(3) Contraction FLEX	(4) Contraction FLEX	(5) Contraction FLEX
D&I rating	-0.071*** (0.014)	-0.065*** (0.014)	-0.054*** (0.014)	-0.053*** (0.015)	-0.039*** (0.015)
Labor stock		0.054** (0.023)	0.027 (0.018)	0.020 (0.019)	0.014 (0.018)
Fixed assets ratio		-0.034* (0.019)	-0.119*** (0.040)	-0.116*** (0.043)	-0.107*** (0.034)
Size		-0.025 (0.022)	0.061** (0.028)	0.047* (0.028)	0.037 (0.029)
Inst. ownership					0.019 (0.022)
Inst. ownership HHI					-0.022 (0.020)
Inside ownership					0.003 (0.017)
_cons	-0.009 (0.023)	-0.008 (0.023)	-0.008 (0.021)	-0.005 (0.021)	-0.030 (0.021)
Observations	9121	8903	8836	8538	7157
R-squared	0.005	0.011	0.142	0.220	0.243
Industry*year FE	No	No	Yes	Yes	Yes
State*year FE	No	No	No	Yes	Yes

Standard errors are in parentheses

*** $p < .01$, ** $p < .05$, * $p < .1$

Table 5: Corporate flexibility after *Vance v. Ball State University* Ruling.

This table examines the effect of a plausibly exogenous shock to Diversity & Inclusion (D&I) on firms' D&I ratings and flexibility metrics following the U.S. Supreme Court decision on June 24 2013 in *Vance v. Ball State University*. The analysis uses a standard difference-in-differences design in which treated firms are headquartered in states covered by the Second, Fourth, and Ninth Circuit Courts, while control firms are headquartered in states covered by the First, Seventh, and Eighth Circuits. The post indicator equals one for years 2013 and later and zero for the pre-treatment year, with the sample restricted to an event window from one year before to five years after the ruling. Flexibility (FLEX) is measured over rolling three-year windows as the standard deviation of log sales growth scaled by the range of operating costs relative to sales. All regressions include firm and year fixed effects, and standard errors are two-way clustered at the firm and state levels. Detailed variable definitions are provided in Appendix B.

	(1) D&I rating	(2) D&I rating	(3) FLEX	(4) FLEX
Treat * Post	-0.104* (0.057)		0.155*** (0.054)	
Treat * Post(t)		-0.234*** (0.084)		0.127 (0.077)
Treat * Post(t+1)		-0.072 (0.075)		0.089 (0.080)
Treat * Post(t+2)		-0.036 (0.057)		0.165* (0.096)
Treat * Post(t+3)		-0.071 (0.059)		0.220*** (0.069)
Treat * Post(t+4)		-0.068 (0.056)		0.236** (0.094)
Treat * Post(t+5)		-0.124* (0.069)		0.110* (0.064)
_cons	0.069** (0.032)	0.066** (0.031)	-0.110*** (0.030)	-0.111*** (0.030)
Observations	6189	6189	6155	6155
R-squared	0.601	0.602	0.419	0.420
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Year FE	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 we 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*** (.004)	.086*** (.028)	.079*** (.028)	.08*** (.029)	.082*** (.029)	
exposure * lagged D&I rating		-.032*** (.008)	-.028*** (.008)	-.027*** (.008)	-.028*** (.009)	-.02** (.009)
exposure * lagged Leverage			-.009** (.004)	-.003 (.005)	-.004 (.005)	0 (.005)
exposure * lagged Cash/assets			-.004 (.007)	-.009 (.008)	-.009 (.008)	-.005 (.007)
exposure * lagged Size			.01** (.004)	.01** (.005)	.008 (.005)	.014** (.006)
exposure * lagged Tobin's Q				.002 (.005)	.002 (.005)	.001 (.005)
exposure * lagged Inst. ownership (%)				-.004 (.005)	-.005 (.005)	0 (.005)
exposure * lagged Inside ownership				-.001 (.004)	-.001 (.004)	-.003 (.004)
exposure * lagged Fixed assets ratio				-.009* (.005)	-.008* (.005)	-.002 (.005)
exposure * lagged Advertising/sales					-.006 (.005)	-.006 (.006)
exposure * lagged Labor stock					-.004 (.008)	-.006 (.008)
exposure * lagged H1-B reliance					.011 (.008)	.014 (.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 we 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*** (.011)	.004 (.081)	-.058 (.084)	-.051 (.087)	-.036 (.088)	
exposure * lagged D&I rating		-.019 (.023)	.006 (.023)	.005 (.024)	.005 (.024)	.015 (.024)
exposure * lagged Leverage			-.003 (.012)	-.009 (.012)	-.013 (.012)	.004 (.012)
exposure * lagged Cash/assets			.002 (.02)	-.003 (.02)	-.004 (.02)	.006 (.019)
exposure * lagged Size			-.006 (.013)	-.014 (.013)	-.017 (.014)	-.013 (.015)
exposure * lagged Tobin's Q				.015 (.011)	.016 (.011)	.013 (.011)
exposure * lagged Inst. ownership (%)				-.039*** (.015)	-.04*** (.014)	-.038*** (.014)
exposure * lagged Inside ownership				-.01 (.016)	-.009 (.015)	-.013 (.013)
exposure * lagged Fixed assets ratio				-.004 (.013)	-.005 (.013)	.014 (.017)
exposure * lagged Advertising/sales					.016 (.014)	.028* (.016)
exposure * lagged Labor stock					0 (.022)	-.006 (.024)
exposure * lagged H1-B reliance					-.013 (.017)	-.014 (.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*** (.003)	-.023 (.017)	-.011 (.019)	-.022 (.02)	-.016 (.02)	
exposure * lagged D&I rating		.009* (.005)	.005 (.005)	.008 (.006)	.006 (.006)	.003 (.006)
exposure * lagged Leverage			0 (.003)	.001 (.003)	.001 (.003)	-.001 (.003)
exposure * lagged Cash/assets			.006 (.005)	.014** (.006)	.012** (.006)	.008 (.006)
exposure * lagged Size			.004 (.003)	.003 (.004)	.003 (.004)	-.002 (.004)
exposure * lagged Tobin's Q				-.005** (.003)	-.005** (.003)	-.005* (.003)
exposure * lagged Inst. ownership (%)				.001 (.003)	.001 (.003)	-.001 (.003)
exposure * lagged Inside ownership				0 (.003)	0 (.003)	.001 (.003)
exposure * lagged Fixed assets ratio				.008** (.003)	.008** (.003)	.002 (.004)
exposure * lagged Advertising/sales					.002 (.004)	0 (.004)
exposure * lagged Labor stock					.001 (.006)	.004 (.006)
exposure * lagged H1-B reliance					.003 (.004)	.002 (.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*** (.012)	.206** (.084)	.177** (.085)	.234** (.094)	.258*** (.095)	
exposure * lagged D&I rating		-.075*** (.024)	-.063*** (.024)	-.074*** (.027)	-.08*** (.027)	-.057** (.026)
exposure * lagged Leverage			-.003 (.013)	.005 (.016)	.001 (.016)	.017 (.014)
exposure * lagged Cash/assets			0 (.017)	-.008 (.021)	-.011 (.021)	.008 (.02)
exposure * lagged Size			-.006 (.014)	-.019 (.017)	-.018 (.017)	-.003 (.017)
exposure * lagged Tobin's Q				.017 (.011)	.016 (.011)	.018 (.011)
exposure * lagged Inst. ownership (%)				-.023 (.016)	-.024 (.016)	-.015 (.015)
exposure * lagged Inside ownership				.008 (.013)	.008 (.012)	-.001 (.012)
exposure * lagged Fixed assets ratio				-.01 (.015)	-.009 (.015)	.02 (.019)
exposure * lagged Advertising/sales					.024 (.015)	.03 (.019)
exposure * lagged Labor stock					.002 (.024)	-.001 (.025)
exposure * lagged H1-B reliance					.021 (.014)	.031* (.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 (4.958)	94.598** (39.176)	110.719*** (40.462)	103.343** (41.378)	101.783** (41.545)	
exposure * lagged D&I rating		-26.22** (11.252)	-29.634** (11.503)	-26.528** (11.604)	-26.999** (11.627)	-20.657* (11.851)
exposure * lagged Leverage			-7.201 (5.929)	-2.986 (7.115)	-2.518 (7.036)	-5.423 (7.48)
exposure * lagged Cash/assets			7.968 (8.947)	-1.614 (10.471)	-1.451 (10.516)	5.022 (9.777)
exposure * lagged Size			18.636*** (6.384)	8.97 (7.233)	9.627 (7.753)	12.414* (7.375)
exposure * lagged Tobin's Q				7.831 (6.023)	7.474 (6.138)	4.399 (6.113)
exposure * lagged Inst. ownership (%)				7.282 (7.114)	6.758 (7.214)	11.116 (6.881)
exposure * lagged Inside ownership				8.789 (11.064)	8.179 (10.944)	10.7 (10.497)
exposure * lagged Fixed assets ratio				.257 (6.206)	.91 (6.213)	8.283 (6.617)
exposure * lagged Advertising/sales					-3.676 (6.809)	-5.33 (6.726)
exposure * lagged Labor stock					-.807 (6.927)	.064 (7.559)
exposure * lagged H1-B reliance					17.287* (10.163)	23.469* (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

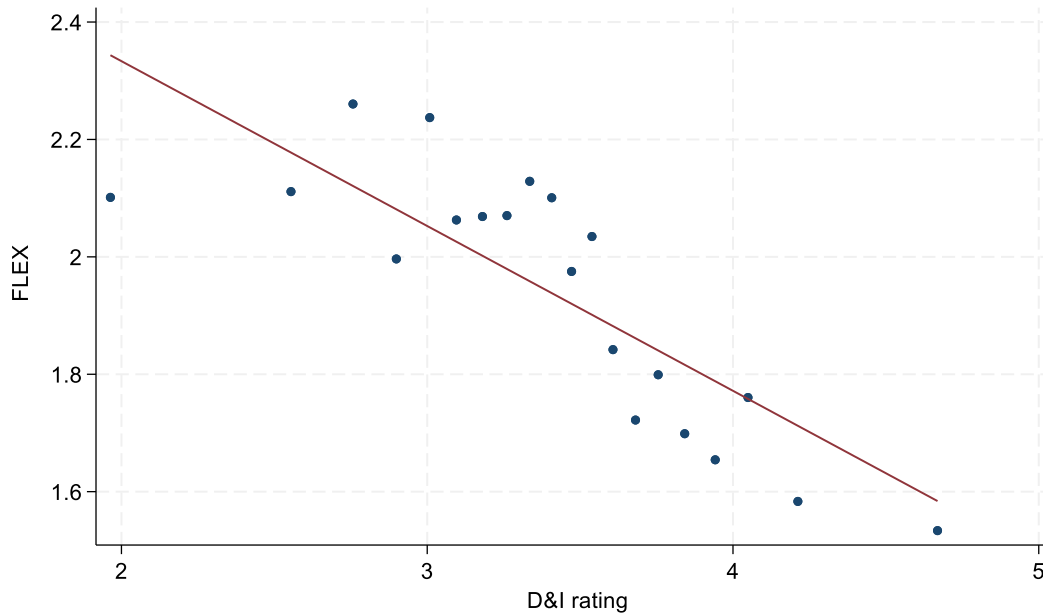
Robust standard errors are in parentheses

*** $p < .01$, ** $p < .05$, * $p < .1$

Figure 1: D&I and Corporate Flexibility.

Panel A plots a bin scatter of corporate flexibility (FLEX) against the D&I rating using the same panel data structure in the baseline Table 3, controlling for labor stock, fixed assets, and firm size, illustrating the conditional relationship without absorbing high-dimensional fixed effects. Panel B plots a bin scatter of FLEX against the D&I rating after residualizing both variables with respect to the same controls and absorbing industry-by-year (FF48×Year) and state-by-year fixed effects.

Panel A: Unconditional bin scatter plot with controls



Panel B: Residualized bin scatter with industry-year and state-year fixed effects

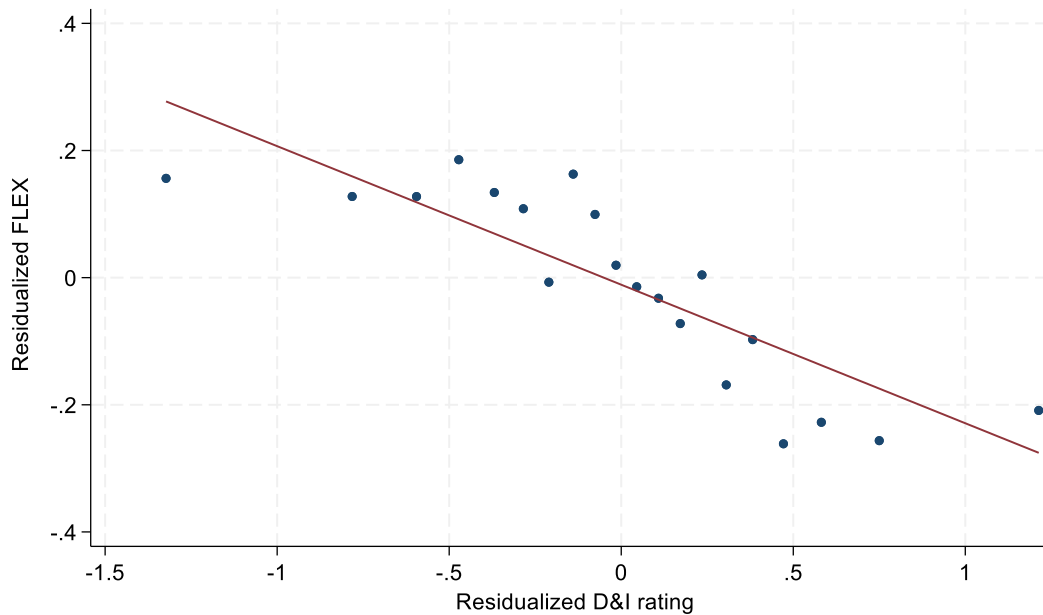
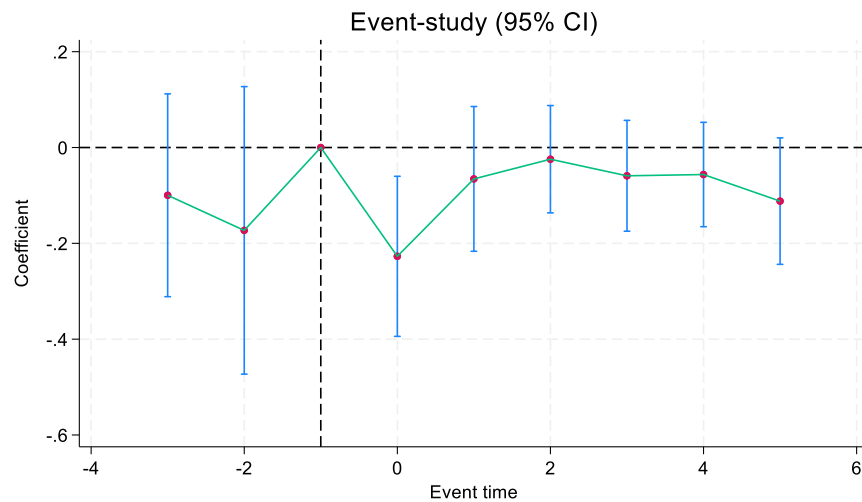


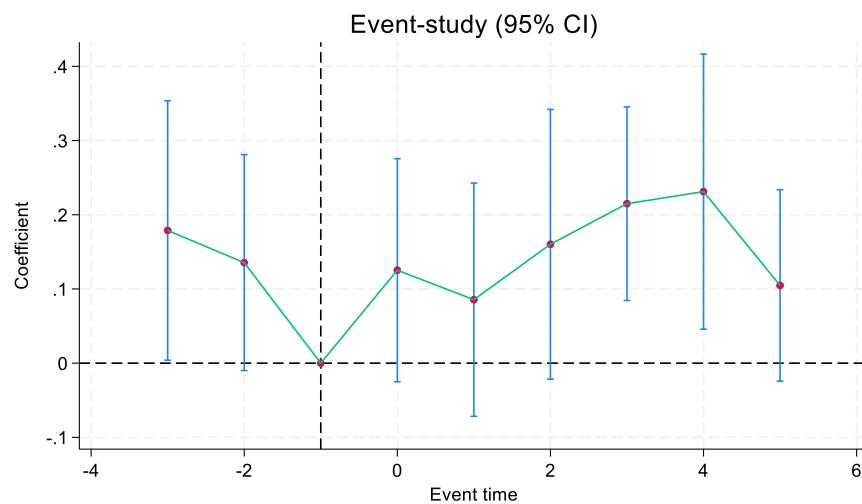
Figure 2: Graphs of D&I and Flexibility around the D&I-related court ruling

This figure presents event-study difference-in-differences estimates around the 2013 *Vance v. Ball State University* Supreme Court ruling. The figure plots coefficient estimates and 95% confidence intervals for the interactions between the treatment indicator and event-time indicators relative to the ruling year. Event time is measured in years relative to 2013, with the year immediately preceding the ruling ($t-1$, 2012) omitted and serving as the reference period. The treatment indicator equals one for firms headquartered in states covered by the Second, Fourth, and Ninth Circuit Courts and zero for firms headquartered in states covered by the First, Seventh, and Eighth Circuits. The dependent variables are the D&I rating and firm-level flexibility (FLEX), where FLEX is measured over rolling three-year windows ($t-2$ to t) as the standard deviation of log sales growth scaled by the range of operating costs relative to sales. All regressions include firm and year fixed effects. Standard errors are two-way clustered at the firm and state levels.

Panel A: D&I rating around the ruling



Panel B: Corporate flexibility (FLEX) around the ruling



INTERNET 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 we 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

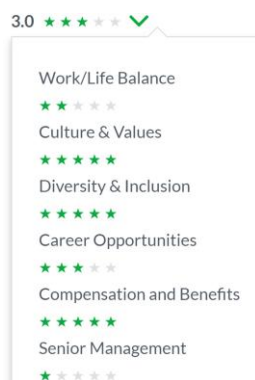
It is LGBTQIA+ friendly, they don't discriminate based on appearances (tattoos and colored hair are welcomed), livable wage, beautiful team culture

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. We 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 we list the detailed description of the variables used in our 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 our own model.

Industry D&I: the average of D&I rating across reviews in firms outside of our 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 benefits 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's 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 our main sample, we start with the list of employers on Glassdoor as of July 2020, each represented by an ID created by Glassdoor (Glassdoor ID). First, we 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 we remove the companies or subsidiaries with a foreign headquarter, e.g., Mumbai (India), arriving at 7,851 unique Glassdoor IDs.

Next, we 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, we 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. We 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), we 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, we manually verify and match each Glassdoor ID with a company in the Compustat database for the 2008-2020 period.¹⁴ 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.

¹⁴ For each Glassdoor ID, we 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, we 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, we find the corresponding latest CUSIP for the company's stock. If the company changed its CUSIP during the sample period, we 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, we drop the company from our sample. However, if the company was acquired or merged with another in 2019 or later, we 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, we assign the CUSIP of the parent firm to the subsidiary. Finally, we drop those Glassdoor IDs associated with a subsidiary of a foreign corporation from our sample.

IA2: Ordinal BERT

The baseline specification of predicting the D&I rating given review texts with BERT treats the five review categories as nominal, optimizing a standard cross-entropy loss over classes 1–5. While effective, this approach does not explicitly encode the fact that misclassifying a 4 as a 5 is less severe than misclassifying it as a 1. The machine learning literature suggests that leveraging this ordinal structure can, in principle, improve performance. Gutierrez et al. (2016) survey a range of methods and report that approaches designed to incorporate ordering information often yield predictions that are closer to the true values along the ordinal scale. At the same time, the evidence is mixed: Ben-David, Sterling, and Tran (2009) compare ordinal and non-ordinal classifiers across many real-world datasets and find that ordinal methods do not consistently dominate. This mixed evidence suggests that the extent of improvement, if any, is likely setting specific.

To explore this, we implemented an ordinal variant of the BERT model by modifying the training objective to penalize errors more heavily when predictions are farther from the true category. Following Sheshadri, Vijjini, and Kharbanda (2021), we optimized a combined loss: cross-entropy plus α (hyper-parameter) times the mean absolute error (MAE) between the predicted and actual categories. The MAE term explicitly encodes ordinal distance, and the weighting parameter α governs the strength of this penalty relative to cross-entropy. Except for the loss, the architecture, tokenizer, data preprocessing, sample, and all other hyperparameters were held fixed to those used in the baseline model with full input data.

To select α , we performed 3-fold cross-validation on a random 10% subsample of the training data. This design balances methodological rigor with computational feasibility: a single full model run requires approximately 15–30 hours on an AWS GPU, and cross-validation multiplies this cost. After choosing α , we retrained the model on the full training set and evaluated on the held-out test set using the same data split as the baseline.

Performance was assessed using two complementary metrics. Accuracy is the conventional metric for multi-class classification but treats all misclassifications equally. To reflect ordinal consistency, we also report correlation with the true labels, which increases when predictions are closer to the correct category and decreases when errors are further away. This dual evaluation allows us to assess performance in both the standard classification sense and the ordinal sense.

The two approaches perform almost identically. On the hold-out testing sample, the baseline model achieved accuracy = 0.6426 and correlation = 0.7713 with the true labels. The ordinal-loss

model achieved accuracy = 0.6424 and correlation = 0.7721. The two sets of predictions are themselves highly correlated (0.9826). Thus, while the ordinal-loss model slightly increases the correlation metric, it also very slightly reduces accuracy, yielding no clear improvement overall.

Given the very large training set, it is plausible that the baseline model already internalizes much of the ordinal structure directly from data, even without explicit ordinal constraints. In light of the effectively indistinguishable out-of-sample performance, the baseline cross-entropy model is retained as the main specification on grounds of simplicity, replicability, and computational efficiency.

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 our sample, and compares that with the distribution of other ratings on Glassdoor.

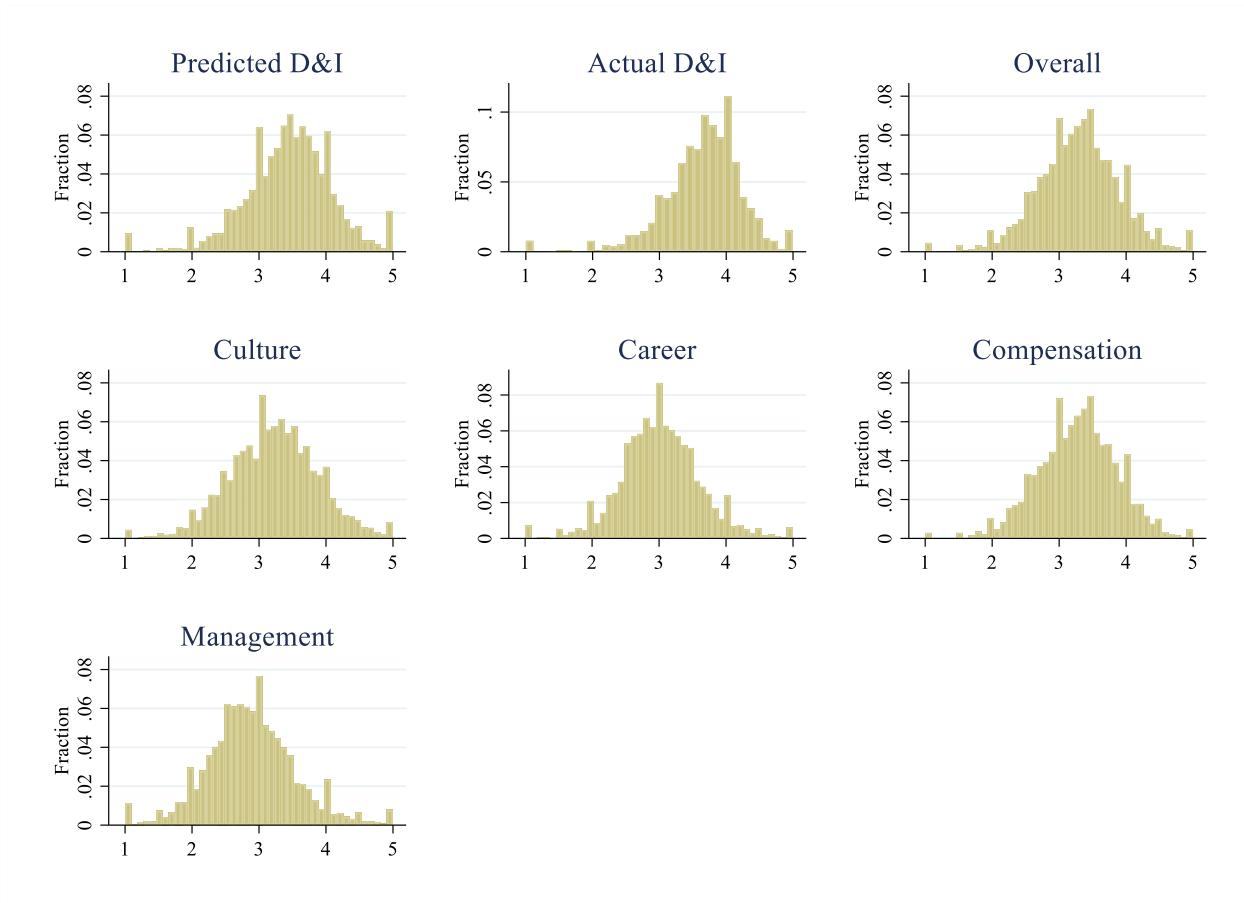


Figure IA2: Trends in D&I

This figure shows the evolution of our D&I measure for the average firm in our 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.

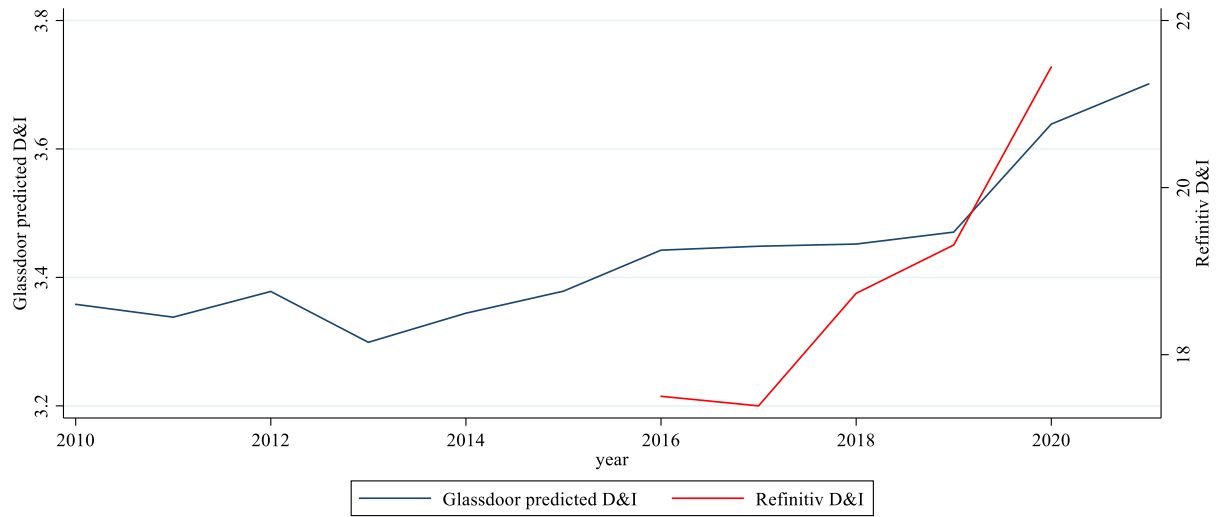
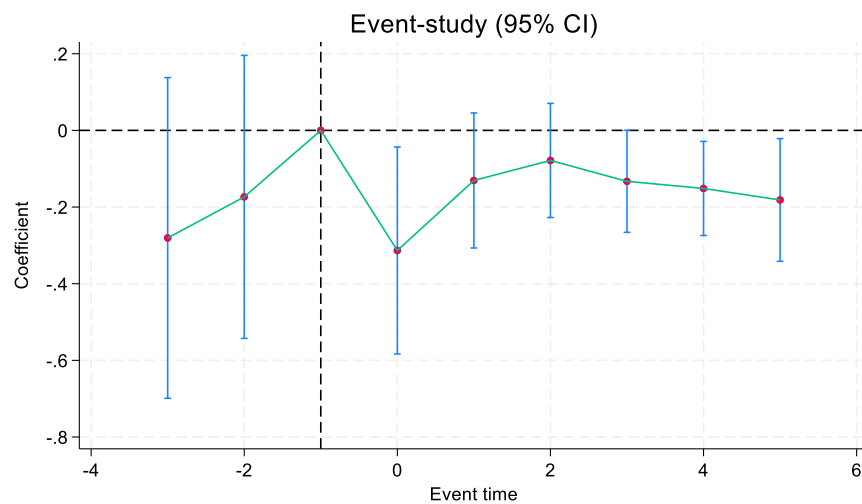


Figure IA3: D&I and Flexibility around the court ruling – matched sample

This figure presents event-study difference-in-differences estimates around the 2013 *Vance v. Ball State University* Supreme Court ruling, similar to Figure 2, but using a matched sample. The figure plots coefficient estimates and 95% confidence intervals for interactions between the treatment indicator and event-time indicators relative to the ruling year, with the year immediately preceding the ruling ($t-1$, 2012) omitted as the reference period. The dependent variables are the D&I rating and firm-level flexibility (FLEX), measured over rolling three-year windows ($t-2$ to t). The matched sample is constructed using one-to-one nearest-neighbor matching based on changes in FLEX in the two and three years prior to the ruling, measured relative to 2012. All regressions include firm, year, and matched-pair fixed effects. Standard errors are two-way clustered at the firm and state levels.

Panel A: D&I rating around the ruling



Panel B: Corporate flexibility (FLEX) around the ruling

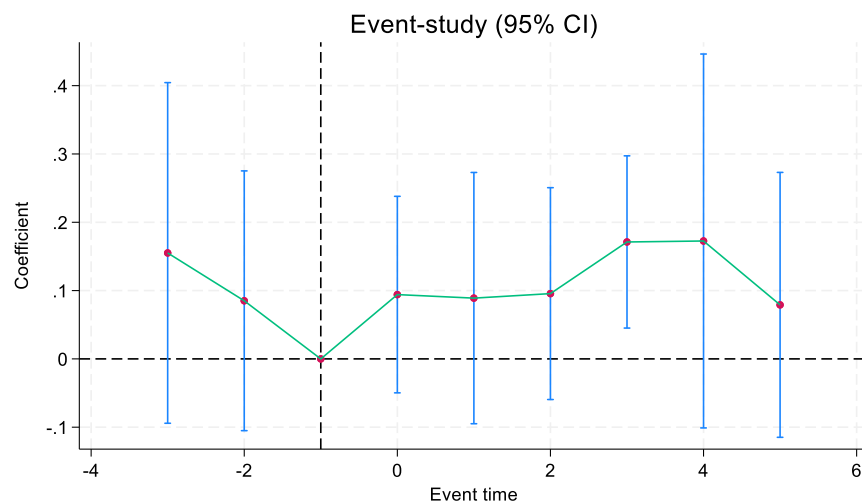


Table IA1: Summary statistics at the review level

This appendix shows the summary statistics of 10.4 million reviews (Panel A), and the correlations among Glassdoor ratings at the review level (Panel B).

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

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. We follow Acikalin et al. (2022)’s methodology to construct this table (see their Table 4), except the filtering criteria that we keep only words appearing in at least 0.1% of the reviews.¹⁵

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

¹⁵ Following the same filter as in Acikalin et al. is too restrictive as it would cut our corpus from 55,000 words to only 267 words. The results remain similar when we 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 our 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*** (.009)	.097*** (.009)	.112*** (.024)
ROA				.044 (.159)	.044 (.159)	.029 (.134)
Sales growth				.128** (.06)	.128** (.06)	.072 (.053)
Tobin's q				.057*** (.01)	.057*** (.01)	.034*** (.008)
Total debt/assets				-.338*** (.064)	-.338*** (.064)	-.08 (.084)
cash/assets				.377*** (.093)	.377*** (.093)	-.011 (.106)
Labor union coverage				-.049 (.308)	-.049 (.308)	-.41 (.311)
Entrenchment index				.013 (.015)	.013 (.015)	-.019 (.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*** (.009)	3.437*** (.009)	3.438*** (0)	2.494*** (.109)	2.494*** (.109)	3.18*** (.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

*** $p < .01$, ** $p < .05$, * $p < .1$

Table IA5: Validating the D&I measure over longer horizons

This table presents tests to validate the predictive power of our D&I measure at the firm-year level over longer horizons. Specifically, we replicate the logit models from Panel A of Table 2, where the dependent variable is an indicator of whether a firm is in *Fortune's Best Workplaces for Diversity* list (Best Diversity list), but shift the prediction window to two and three years ahead. All independent variables—including the D&I measure, other Glassdoor ratings (overall rating, culture rating) are averaged to the firm-year level across reviews and measured prior to the prediction outcome. The Refinitiv D&I rating is the average of diversity and inclusion scores provided by Refinitiv. Control variables include firm size, ROA, and leverage (total debts over assets). Industry classifications are based on Fama-French 48 industries. All models include an intercept. Standard errors are clustered at the firm level.

Panel A: Prediction two years ahead

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D&I rating	1.789*** (.151)	2.102*** (.212)	2.049*** (.27)	2.838*** (.336)	2.637*** (.441)	3.372*** (.551)	2.043*** (.469)	1.688*** (.601)
Lagged Best Diversity			4.276*** (.392)		3.617*** (.404)			
Refinitiv D&I						.05*** (.009)		
Overall rating							.063 (.439)	
Culture rating								.388 (.55)
Observations	9691	5428	3268	5142	3098	1161	5428	5425
Pseudo R ²	.095	.145	.335	.267	.397	.28	.145	.146
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

Panel B: Prediction three years ahead

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D&I rating	1.479*** (.134)	1.802*** (.185)	1.802*** (.266)	2.371*** (.299)	2.158*** (.44)	2.023*** (.575)	1.49*** (.367)	1.746*** (.343)
Lagged Best Diversity			4.036*** (.498)		3.693*** (.558)			
Refinitiv D&I						.045*** (.013)		
Overall rating							.338 (.348)	
Culture rating								.083 (.311)
Observations	9525	5354	2189	5052	1961	428	5354	5335
Pseudo R ²	.074	.125	.249	.243	.323	.198	.126	.127
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

*** $p < .01$, ** $p < .05$, * $p < .1$

Table IA6: Robustness to baseline results in Table 3

This table examines the robustness of the cross-sectional relationship between corporate flexibility (FLEX) and Diversity & Inclusion (D&I) using the same baseline specification as Table 3, with each panel modifying one element of the baseline. Panel A excludes firms with fewer than ten years of operating cost margin data. Panels B and C limit the sample to firm-years with at least five and ten Glassdoor reviews, respectively. Panel D adds the Bebchuk, Cohen, and Ferrell (2009) entrenchment index as a control, Panel E includes firms' D&I policy information from Refinitiv as a control, Panel F adds board and top management diversity measures as controls, and Panel G includes cash and debt ratios as controls. Panel H incorporates lead-lag specifications (regressing flexibility estimated during year t to t+5 period on D&I at year t), and Panel I re-estimates FLEX using quarterly data. Across all panels, continuous variables are standardized, ratio variables are winsorized at the 1st and 99th percentiles, and standard errors are clustered by firm.

Panel A: Excluding firms with fewer than ten years of cost margin data

	(1) FLEX	(2) FLEX	(3) FLEX	(4) FLEX	(5) FLEX
D&I rating	-0.067*** (0.014)	-0.065*** (0.015)	-0.057*** (0.015)	-0.053*** (0.016)	-0.058*** (0.017)
Labor stock		0.081*** (0.025)	0.058*** (0.022)	0.047** (0.021)	0.035* (0.020)
Fixed assets ratio		-0.043* (0.024)	-0.164*** (0.055)	-0.161*** (0.057)	-0.156*** (0.047)
Size		0.011 (0.031)	0.116*** (0.035)	0.095*** (0.035)	0.108*** (0.038)
Inst. ownership					-0.026 (0.028)
Inst. ownership HHI					-0.046* (0.023)
Inside ownership					-0.018 (0.017)
_cons	-0.002 (0.028)	0.007 (0.028)	0.007 (0.026)	0.009 (0.026)	0.006 (0.027)
Observations	15902	14241	14194	13741	11578
R-squared	0.005	0.012	0.182	0.234	0.247
Industry*year FE	No	No	Yes	Yes	Yes
State*year FE	No	No	No	Yes	Yes

Standard errors are in parentheses

*** $p < .01$, ** $p < .05$, * $p < .1$

Panel B: Restricting the sample to firm-years with at least five Glassdoor reviews

	(1) FLEX	(2) FLEX	(3) FLEX	(4) FLEX	(5) FLEX
D&I rating	-0.095*** (0.015)	-0.084*** (0.016)	-0.066*** (0.016)	-0.060*** (0.018)	-0.053*** (0.017)
Labor stock		0.069*** (0.020)	0.044*** (0.016)	0.036** (0.016)	0.030* (0.017)
Fixed assets ratio		-0.046** (0.020)	-0.159*** (0.043)	-0.157*** (0.045)	-0.140*** (0.038)
Size		0.017 (0.026)	0.102*** (0.031)	0.085*** (0.032)	0.078** (0.032)
Inst. ownership					-0.010 (0.023)
Inst. ownership HHI					-0.062** (0.025)
Inside ownership					-0.019 (0.014)
_cons	-0.004 (0.024)	-0.025 (0.023)	-0.026 (0.022)	-0.023 (0.022)	-0.044** (0.022)
Observations	18484	15026	14978	14527	12134
R-squared	0.009	0.014	0.171	0.225	0.228
Industry*year FE	No	No	Yes	Yes	Yes
State*year FE	No	No	No	Yes	Yes

Standard errors are in parentheses

*** $p < .01$, ** $p < .05$, * $p < .1$

Panel C: Restricting the sample to firm-years with at least ten Glassdoor reviews

	(1) FLEX	(2) FLEX	(3) FLEX	(4) FLEX	(5) FLEX
D&I rating	-0.104*** (0.016)	-0.090*** (0.017)	-0.073*** (0.018)	-0.068*** (0.019)	-0.060*** (0.018)
Labor stock		0.065*** (0.020)	0.041** (0.016)	0.033** (0.016)	0.029* (0.017)
Fixed assets ratio		-0.047** (0.020)	-0.164*** (0.044)	-0.161*** (0.046)	-0.143*** (0.039)
Size		0.014 (0.027)	0.102*** (0.032)	0.087*** (0.033)	0.079** (0.033)
Inst. ownership					-0.010 (0.025)
Inst. ownership HHI					-0.062** (0.026)
Inside ownership					-0.019 (0.014)
_cons	-0.003 (0.025)	-0.024 (0.024)	-0.026 (0.022)	-0.023 (0.022)	-0.043* (0.022)
Observations	16472	13426	13374	12961	10847
R-squared	0.011	0.015	0.170	0.228	0.231
Industry*year FE	No	No	Yes	Yes	Yes
State*year FE	No	No	No	Yes	Yes

Standard errors are in parentheses

*** $p < .01$, ** $p < .05$, * $p < .1$

Panel D: Controlling for Bebchuk, Cohen, and Ferrell (2009) entrenchment index

	(1) FLEX	(2) FLEX	(3) FLEX	(4) FLEX	(5) FLEX
D&I rating	-0.100*** (0.019)	-0.089*** (0.022)	-0.088*** (0.022)	-0.088*** (0.023)	-0.070*** (0.021)
entrenchment index	0.025 (0.019)	0.019 (0.019)	0.068*** (0.021)	0.057*** (0.022)	0.059** (0.023)
Labor stock		0.070*** (0.024)	0.044** (0.018)	0.039** (0.018)	0.046** (0.020)
Fixed assets ratio		-0.070*** (0.023)	-0.190*** (0.051)	-0.177*** (0.053)	-0.162*** (0.049)
Size		-0.002 (0.035)	0.140*** (0.045)	0.126*** (0.044)	0.129*** (0.044)
Inst. ownership					0.033 (0.036)
Inst. ownership HHI					0.020 (0.050)
Inside ownership					-0.021 (0.014)
_cons	-0.003 (0.029)	-0.012 (0.026)	-0.049** (0.024)	-0.046* (0.024)	-0.058** (0.026)
Observations	12442	10695	10648	10556	9250
R-squared	0.009	0.019	0.207	0.270	0.279
Industry*year FE	No	No	Yes	Yes	Yes
State*year FE	No	No	No	Yes	Yes

Standard errors are in parentheses

*** $p < .01$, ** $p < .05$, * $p < .1$

Panel E: Controlling for corporate D&I policies from Refinitiv

	(1) FLEX	(2) FLEX	(3) FLEX	(4) FLEX	(5) FLEX
D&I rating	-0.133*** (0.023)	-0.133*** (0.023)	-0.089*** (0.021)	-0.070*** (0.022)	-0.071*** (0.022)
D&I rating - Refinitiv	0.041 (0.026)	0.054* (0.030)	-0.000 (0.031)	-0.001 (0.031)	0.002 (0.034)
Labor stock		0.096*** (0.034)	0.066*** (0.022)	0.058** (0.023)	0.056** (0.024)
Fixed assets ratio		-0.047** (0.021)	-0.194*** (0.046)	-0.177*** (0.048)	-0.156*** (0.044)
Size		-0.011 (0.038)	0.133*** (0.047)	0.109** (0.043)	0.109** (0.046)
Inst. ownership					-0.010 (0.029)
Inst. ownership HHI					-0.051 (0.032)
Inside ownership					-0.017 (0.016)
_cons	-0.037 (0.028)	-0.018 (0.028)	-0.055** (0.025)	-0.049* (0.026)	-0.061** (0.025)
Observations	4761	4642	4616	4546	4240
R-squared	0.013	0.021	0.183	0.250	0.259
Industry*year FE	No	No	Yes	Yes	Yes
State*year FE	No	No	No	Yes	Yes

Standard errors are in parentheses

*** $p < .01$, ** $p < .05$, * $p < .1$

Panel F: Controlling for board and top management diversity

	(1) FLEX	(2) FLEX	(3) FLEX	(4) FLEX	(5) FLEX
D&I rating	-0.073*** (0.013)	-0.073*** (0.014)	-0.059*** (0.013)	-0.055*** (0.014)	-0.057*** (0.014)
Board nationality mix	-0.017 (0.019)	-0.018 (0.019)	-0.022 (0.019)	-0.016 (0.018)	-0.017 (0.018)
Board gender ratio	-0.028 (0.019)	-0.028 (0.019)	-0.030 (0.023)	-0.024 (0.024)	-0.026 (0.025)
Board age diversity	-0.042** (0.021)	-0.051** (0.021)	-0.037* (0.021)	-0.031 (0.021)	-0.036* (0.019)
Board diversity: firm tenure	-0.021 (0.018)	-0.018 (0.019)	-0.041** (0.020)	-0.038* (0.020)	-0.022 (0.020)
Board diversity: qualifications	0.014 (0.022)	0.012 (0.022)	0.008 (0.022)	0.008 (0.023)	0.009 (0.022)
Board: network size	-0.025 (0.022)	-0.019 (0.026)	-0.046 (0.029)	-0.018 (0.030)	-0.026 (0.021)
Labor stock		0.074*** (0.023)	0.051** (0.021)	0.043** (0.020)	0.035* (0.019)
Fixed assets ratio		-0.040* (0.022)	-0.157*** (0.044)	-0.146*** (0.045)	-0.136*** (0.037)
Size		-0.000 (0.031)	0.111*** (0.037)	0.069* (0.037)	0.081** (0.035)
Inst. ownership					-0.011 (0.023)
Inst. ownership HHI					-0.054** (0.024)
Inside ownership					-0.014 (0.013)
_cons	-0.031 (0.024)	-0.024 (0.024)	-0.028 (0.022)	-0.025 (0.022)	-0.047** (0.021)
Observations	15321	14605	14568	14430	13305
R-squared	0.010	0.017	0.166	0.217	0.224
Industry*year FE	No	No	Yes	Yes	Yes
State*year FE	No	No	No	Yes	Yes

Standard errors are in parentheses

*** $p < .01$, ** $p < .05$, * $p < .1$

Panel G: Controlling for cash and debt ratios

	(1) FLEX	(2) FLEX	(3) FLEX	(4) FLEX	(5) FLEX
D&I rating	-0.062*** (0.013)	-0.046*** (0.014)	-0.042*** (0.013)	-0.043*** (0.014)	-0.038*** (0.014)
Leverage	0.005 (0.017)	0.032* (0.018)	0.026 (0.019)	0.027 (0.020)	0.025 (0.022)
Cash/assets	-0.120*** (0.018)	-0.162*** (0.022)	-0.121*** (0.023)	-0.106*** (0.023)	-0.113*** (0.026)
Labor stock		0.066*** (0.020)	0.045** (0.017)	0.038** (0.018)	0.028 (0.018)
Fixed assets ratio		-0.086*** (0.021)	-0.171*** (0.042)	-0.164*** (0.044)	-0.147*** (0.037)
Size		-0.051* (0.026)	0.054* (0.030)	0.042 (0.031)	0.032 (0.032)
Inst. ownership					-0.008 (0.023)
Inst. ownership HHI					-0.073*** (0.024)
Inside ownership					-0.014 (0.013)
_cons	-0.033 (0.022)	-0.024 (0.022)	-0.026 (0.021)	-0.024 (0.021)	-0.043** (0.021)
Observations	17627	16872	16837	16333	13540
R-squared	0.023	0.040	0.177	0.223	0.228
Industry*year FE	No	No	Yes	Yes	Yes
State*year FE	No	No	No	Yes	Yes

Standard errors are in parentheses

*** $p < .01$, ** $p < .05$, * $p < .1$

Panel H: Allowing for lead-lag relationship

	(1) FLEX	(2) FLEX	(3) FLEX	(4) FLEX	(5) FLEX
D&I rating	-0.042*** (0.016)	-0.036*** (0.012)	-0.033*** (0.011)	-0.030** (0.012)	-0.022** (0.011)
Labor stock		0.047*** (0.014)	0.038*** (0.013)	0.029** (0.013)	0.020* (0.012)
Fixed assets ratio		-0.052*** (0.013)	-0.117*** (0.025)	-0.116*** (0.026)	-0.106*** (0.025)
Size		0.020 (0.017)	0.056*** (0.021)	0.036** (0.018)	0.043** (0.021)
Inst. ownership					0.007 (0.017)
Inst. ownership HHI					0.001 (0.027)
Inside ownership					-0.010 (0.009)
_cons	-0.000 (0.016)	0.006 (0.016)	0.005 (0.015)	0.006 (0.015)	0.005 (0.015)
Observations	15308	13805	13773	13360	11201
R-squared	0.002	0.005	0.105	0.153	0.177
Industry*year FE	No	No	Yes	Yes	Yes
State*year FE	No	No	No	Yes	Yes

Standard errors are in parentheses

*** $p < .01$, ** $p < .05$, * $p < .1$

Panel I: Using quarterly data to measure flexibility

	(1) FLEX	(2) FLEX	(3) FLEX	(4) FLEX	(5) FLEX
D&I rating	-0.055*** (0.015)	-0.054*** (0.016)	-0.036** (0.014)	-0.034** (0.015)	-0.032** (0.016)
Labor stock		0.096*** (0.027)	0.071*** (0.023)	0.060*** (0.022)	0.049** (0.023)
Fixed assets ratio		0.006 (0.022)	-0.134*** (0.038)	-0.131*** (0.041)	-0.128*** (0.041)
Size		0.045* (0.025)	0.128*** (0.027)	0.102*** (0.028)	0.096*** (0.032)
Inst. ownership					-0.020 (0.026)
Inst. ownership HHI					-0.070** (0.031)
Inside ownership					-0.021 (0.018)
_cons	0.004 (0.024)	0.005 (0.024)	0.006 (0.021)	0.005 (0.021)	-0.005 (0.022)
Observations	20159	16594	16558	16106	13384
R-squared	0.003	0.012	0.276	0.311	0.321
Industry*year FE	No	No	Yes	Yes	Yes
State*year FE	No	No	No	Yes	Yes

Standard errors are in parentheses

*** $p < .01$, ** $p < .05$, * $p < .1$

Table IA7: Robustness to uncertainty in the D&I rating

This table examines whether the cross-sectional relationship between corporate flexibility (FLEX) and Diversity & Inclusion (D&I) is sensitive to uncertainty in the machine-learning-based D&I measure. Panel A replaces the baseline measure with a probability-weighted average that uses the full distribution of BERT-predicted label probabilities for each review. Panel B implements a three-layer bootstrap procedure designed to incorporate uncertainty at the model, review, and firm-year levels. Specifically, for each bootstrap replication, firm-years are resampled with replacement; reviews within each firm-year are resampled to match the observed number of reviews; and a simulated D&I label is assigned to each review by drawing from the BERT-predicted probability vector. These simulated labels are then aggregated into firm-year D&I ratings, merged back into the resampled panel, standardized using means and standard deviations computed from the original baseline sample, and used to re-estimate the same five regression specifications as in Table 3. Each regression is estimated with firm-clustered standard errors, and the bootstrap standard errors are computed as the standard deviation of the coefficient estimates across 1,000 replications. Continuous variables are standardized, ratio variables are winsorized at the 1st and 99th percentiles, and all other sample restrictions follow the baseline.

Panel A: Probability-weighted D&I ratings

	(1) FLEX	(2) FLEX	(3) FLEX	(4) FLEX	(5) FLEX
Probability-weighted D&I rating	-0.053*** (0.011)	-0.044*** (0.012)	-0.083*** (0.017)	-0.081*** (0.018)	-0.073*** (0.017)
Labor stock		0.077*** (0.022)	0.052*** (0.018)	0.043** (0.018)	0.033* (0.018)
Fixed assets ratio		-0.036* (0.020)	-0.144*** (0.042)	-0.142*** (0.044)	-0.125*** (0.037)
Size		0.007 (0.025)	0.098*** (0.029)	0.078*** (0.030)	0.073** (0.031)
Inst. ownership					-0.010 (0.023)
Inst. ownership HHI					-0.057** (0.024)
Inside ownership					-0.019 (0.013)
_cons	-0.003 (0.024)	-0.028 (0.023)	-0.033 (0.021)	-0.029 (0.021)	-0.050** (0.021)
Observations	20879	16872	16837	16333	13540
R-squared	0.003	0.009	0.165	0.215	0.219
Industry*year FE	No	No	Yes	Yes	Yes
State*year FE	No	No	No	Yes	Yes

Standard errors are in parentheses

*** $p < .01$, ** $p < .05$, * $p < .1$

Panel B: Three-layer bootstrap simulation of D&I ratings

	(1) FLEX	(2) FLEX	(3) FLEX	(4) FLEX	(5) FLEX
D&I rating	-0.035*** (0.008)	-0.029*** (0.009)	-0.043*** (0.010)	-0.042*** (0.010)	-0.038*** (0.010)
Labor stock		0.078*** (0.007)	0.053*** (0.006)	0.046*** (0.006)	0.035*** (0.007)
Fixed assets ratio		-0.039*** (0.007)	-0.149*** (0.014)	-0.142*** (0.015)	-0.129*** (0.014)
Size		0.006 (0.008)	0.093*** (0.009)	0.074*** (0.009)	0.068*** (0.010)
Inst. ownership					-0.010 (0.010)
Inst. ownership HHI					-0.052*** (0.012)
Inside ownership					-0.020*** (0.006)
_cons	-0.008 (0.007)	-0.029*** (0.008)	-0.082 (0.068)	-0.070 (0.163)	-0.567*** (0.199)
Observations	20159	16594	16558	16106	13384
R-squared	0.003	0.012	0.276	0.311	0.321
Industry*year FE	No	No	Yes	Yes	Yes
State*year FE	No	No	No	Yes	Yes

Standard errors are in parentheses

*** $p < .01$, ** $p < .05$, * $p < .1$

Table IA8: Controlling for past flexibility

This table tests the robustness of the relationship between corporate flexibility (FLEX) and Diversity & Inclusion (D&I) by adding a control for firms' historical flexibility. We construct a pre-period FLEX measure using the same method as in the baseline and over a window of the same length as the main estimation period. Specifically, the pre-period window is 1996–2008, as opposed to 2009–2021 for the baseline period. The resulting firm-level pre-period FLEX measure (“past FLEX”) is merged into the main panel and included as an additional regressor, after standardization to have zero mean and unit standard deviation. Apart from adding this historical measure, all other aspects of the specification—including controls, fixed effects, sample restrictions, and clustering—follow Table 3.

	(1) FLEX	(2) FLEX	(3) FLEX	(4) FLEX	(5) FLEX
D&I rating	-0.062*** (0.015)	-0.057*** (0.015)	-0.054*** (0.015)	-0.054*** (0.016)	-0.050*** (0.016)
past FLEX	0.376*** (0.051)	0.355*** (0.049)	0.277*** (0.045)	0.271*** (0.044)	0.250*** (0.039)
Labor stock		0.048** (0.020)	0.043** (0.019)	0.039** (0.019)	0.031 (0.019)
Fixed assets ratio		-0.054*** (0.020)	-0.149*** (0.044)	-0.149*** (0.047)	-0.139*** (0.040)
Size		0.001 (0.026)	0.077** (0.031)	0.057* (0.032)	0.059* (0.032)
Inst. ownership					-0.003 (0.026)
Inst. ownership HHI					-0.027 (0.029)
Inside ownership					-0.008 (0.013)
_cons	0.001 (0.025)	-0.020 (0.024)	-0.023 (0.023)	-0.021 (0.024)	-0.039* (0.023)
Observations	16248	13919	13875	13449	11297
R-squared	0.143	0.150	0.254	0.297	0.301
Industry*year FE	No	No	Yes	Yes	Yes
State*year FE	No	No	No	Yes	Yes

Standard errors are in parentheses

*** $p < .01$, ** $p < .05$, * $p < .1$