

ESG Performance and Voluntary ESG Disclosure: Mind the (Gender Pay) Gap*

June Huang[†] and Shirley Lu[‡]

May 2022

Abstract

We study if firms with better ESG performance are more likely to provide voluntary ESG disclosure, an assumption embedded in many ESG ratings. We focus on gender diversity and proxy for performance using a firm's gender pay gap ("GPG") disclosed under a UK disclosure mandate. Contrary to the prediction of traditional disclosure models, before the mandate, we find firms with better GPGs are less likely to voluntarily disclose other information on gender diversity. As a practical implication, these firms have lower social scores from ESG rating companies. We find suggestive evidence that firms with better GPGs do not need to disclose because they already receive workplace diversity awards and that firms with worse GPGs engage in selective disclosure.

Keywords: Gender pay gap, gender diversity, voluntary disclosure, ESG ratings, ESG, corporate social responsibility.

JEL Classification: D80, M14, M41

*We thank Marianne Bertrand, Hans Christensen, Christian Leuz, Lisa Liu, Haresh Sapra, Heather Sarsons (discussant), Abbie Smith, Kate Suslava (discussant), workshop participants at the University of Chicago, and participants at the Chicago-LSE "Economics of Social Sector Organizations" conference, the Stanford GSB "Rising Scholars" conference, and the AAA Annual Meeting for helpful comments and insights. We thank Georgios Tzortzis for excellent research assistance. All errors are our own. A previous version of this paper was circulated under the title, "How Informative are Voluntary CSR disclosures? Evidence from the Gender Pay Gap."

[†]University of Chicago Booth School of Business (jhuang16@chicagobooth.edu)

[‡]Harvard Business School (slu@hbs.edu)

1 Introduction

As more investors make decisions based on environmental, social, and governance (“ESG”) issues, firms increasingly provide voluntary disclosure on ESG (Christensen et al., 2021b). This voluntary disclosure is a key input for ESG ratings, which assign higher scores to firms that provide more voluntary disclosure (e.g., ThomsonReuters, 2013; Sustainalytics, 2017; Li et al., 2021). Spurred by the rising importance of these disclosures, we ask whether firms with better ESG performance are more or less likely to provide voluntary ESG disclosure.

This is ultimately an empirical question. If ESG disclosure is similar to financial disclosure, then traditional financial disclosure models predict firms with better performance are more likely to provide voluntary disclosure (e.g., Verrecchia, 1983). However, this prediction may not apply because ESG disclosure differs from financial disclosure (Christensen et al., 2021b). First, while periodic mandatory reporting plays a confirmatory role for voluntary financial disclosure (Ball et al., 2012), ESG disclosure often takes place without this disciplining mechanism. If ESG disclosure is less credible as a result, then we would not expect to find any association between ESG performance and voluntary ESG disclosure (Crawford and Sobel, 1982). Second, ESG stakeholders have historically used alternative sources of information to learn about firm ESG performance, such as reports from environmental NGOs or websites that aggregate employee reviews. If this alternative information identifies firms with better ESG performance, then these firms may not need to provide their own voluntary ESG disclosure. Instead, firms with worse performance based on alternative information would be more likely to disclose, leading to a negative association between ESG performance and disclosure.

To answer this question, we focus on the ESG issue of gender diversity and exploit a UK mandate (“GPG Mandate”) that required firms with over 250 employees to disclose gender pay gap (“GPG”) data for 2017 and onward.¹ We consider GPG an important component of a firm’s gender diversity, and we examine whether firms with better GPGs, based on 2017 data, are more or less likely to provide voluntary gender diversity disclosure in the annual report released in 2016.² This setting

¹The regulation’s full name is “Equality Act 2010 (Gender Pay Gap Information) Regulations 2017, SI 2017/172”.

²We observe that in the UK, when firms disclose about gender diversity, they tend to do so in their annual reports, and their sustainability reports tend to mimic the gender diversity text in the annual reports. As such, we focus on

provides us with two advantages. First, GPG Mandate strictly defines how firms should calculate GPG data, making the measure comparable across firms. Second, by measuring voluntary disclosure before the mandate, we can examine how firms voluntarily disclose in a typical ESG disclosure environment where there are few mandatory reporting requirements.

We focus on the Financial Times Stock Exchange 350 (“FTSE 350”) at the end of 2016 and hand-collect voluntary gender diversity disclosure.³ These firms are more likely to have sophisticated investor bases that rely on ESG disclosure to evaluate gender diversity performance. For example the Investment Association, a large UK investor group, specifically tracks the gender diversity progress of the FTSE 350 (Kollewe and Butler, 2018). Out of the FTSE 350, we drop firms exempt from the mandate, go public in 2015 or 2016, or mention future GPG regulation in their annual report released in 2016. In our main analysis, we focus on 126 firms where we have GPG data for the consolidated firm.⁴

First, we document how firms voluntarily disclose in their annual reports. On the extensive margin, only 46% of firms provide voluntary disclosure on gender diversity. Disclosures fall into five categories: (i) metrics, (ii) targets, (iii) external commitments, (iv) internal actions, and (v) policy benefits. Metrics are quantitative and include proportions of women at entry and senior levels. Targets are defined by a quantitative diversity goal and year. Internal actions include unconscious bias training and programs encouraging women through the corporate pipeline. External commitments include membership with external gender diversity organizations such as the 30% Club. Finally, policy benefits are benefits given to employees, such as parental leave and flexible working programs. The average firm in the sample discloses 1.6 categories, and the most commonly disclosed categories are internal actions and metrics.

Next, we turn to our main question: are firms with better GPGs more or less likely to provide voluntary gender diversity disclosure? We define our main independent variable of interest as *Good GPG* = $1 - GPG$, which transforms a smaller GPG into a higher *Good GPG*. For example, the

voluntary disclosure in annual reports.

³The FTSE 350 is made up of the 350 largest firms by market capitalization whose primary listing is on the London Stock Exchange.

⁴GPG Mandate applies at the hiring entity-level, which is a subset of the consolidated firm. For this reason, we do not have consolidated firm-level GPG for some firms. We discuss the sample in more detail in section 3.1.

average gender pay gap in our sample is 20%, which translates to a *Good GPG* of 0.80. We regress disclosure outcomes on *Good GPG*. We include 2-digit NAICS industry fixed effects to conduct the analysis within-industry because firms in the same industry share similar labor supply conditions that may impact GPG (Cook et al., 2021); investors often evaluate firm ESG performance within-industry for the same reason. While traditional financial disclosure models predict firms with better GPGs are more likely to disclose (e.g., Verrecchia, 1983), we find the opposite: firms with better GPGs are significantly less likely to voluntarily disclose. Specifically, having a 10% higher *Good GPG* is associated with a 17% lower likelihood of disclosure. This result holds across a number of robustness tests, such as defining industry using 3-digit NAICS.

Our result has a practical implication for ESG ratings: if firms with more disclosure are given higher ESG ratings, then firms with worse GPGs will have higher ESG ratings. Many investors rely on ESG ratings to evaluate a firm's ESG performance (Hartzmark and Sussman, 2019); however, recent work finds low correlation across ESG ratings from different rating providers (Chatterji et al., 2016; Berg et al., 2020). One source of variation comes from providers placing different weights on firm disclosure. For example, some ratings penalize firms for not providing ESG disclosure on a particular topic (Li et al., 2021). We examine social scores from three popular ESG rating providers: Asset4 from Thomson Reuters, Sustainalytics from Morningstar, and MSCI. We find that firms with worse GPGs have higher Asset4 social scores, and GPG is not meaningfully related to Sustainalytics or MSCI social scores in either direction. This result is consistent with Asset4's heavier reliance on voluntary disclosure during our time period: relative to Sustainalytics or MSCI, Asset4's social score correlates more highly with our measures of voluntary disclosure. Consequently, our findings join recent work in advising caution when using ESG ratings to assess ESG performance.

Finally, we attempt to understand why firms with better GPGs are less likely to provide voluntary disclosure. Because stakeholders use alternative information to learn more about ESG issues, we explore if our results are driven by the availability of labor market information. In particular, employees who care about workplace gender diversity can seek information from publications like the *The Times* Top 50 Employers for Women. We find that firms with better GPGs are more likely to receive this award. As a result, these firms may not need to provide additional voluntary

disclosure. Instead, firms with worse performance would be more likely to disclose. Because there are multiple ways to measure ESG performance (Christensen et al., 2021b), firms with worse GPGs may selectively disclose metrics that appear more favorable. Indeed, when firms disclose, the sentiment of their words are more often neutral or positive rather than negative. Consistent with selective disclosure, we find that among firms with worse GPGs, those with more highly paid women are more likely to disclose metrics highlighting women in leadership positions. However, we hesitate to over-interpret these explanations because we acknowledge that our small sample prevents us from ruling out alternative explanations.

Our analysis relies on two assumptions. First, we assume a firm's gender pay gap is an important component of its gender diversity, allowing us to use GPG as a measure of a firm's latent gender diversity performance. We believe this assumption is reasonable because GPG is affected by both gender composition and pay by gender at each job level. Furthermore, the UK Government Equalities Office ("GEO") views GPG reporting as "a valuable tool for assessing levels of equality in the workplace, male and female participation, and how effectively talent is being maximized" (GEO 2019, p. 4). Second, we assume a firm's GPG reported for 2017 is a good proxy for its GPG one year earlier when it released an annual report in 2016. GPG is likely slow-moving because it is shaped by industry-specific labor supply conditions and management's tone at the top. However, one potential concern is that in anticipation of the GPG Mandate, firms change their GPGs between 2016 and 2017. Though we lack reliable GPG data before 2017, we examine firm gender composition from 2015-2017, and we find that gender composition stayed relatively constant. Taken together, these reasons help alleviate the concern that a firm's GPG differs significantly between 2017 and when the firm released the annual report in 2016.

We make several contributions to the literature on voluntary ESG disclosure. First, we document the surprising result that firms with better ESG performance are less likely to provide voluntary ESG disclosure. Prior papers on environmental performance and voluntary environmental disclosure present mixed results. Al-Tuwaijri et al. (2004) and Clarkson et al. (2008) show a positive association between environmental performance and voluntary environmental disclosure; Ingram and Frazier (1980) and Wiseman (1982) show no association; and Patten (2002) shows a negative association.

These papers either use an environmental performance measure that is not comparable across firms or measure voluntary disclosure that firms made knowing it would be confirmed ex-post.⁵ By contrast, we use a comparable measure of gender diversity performance, and we measure voluntary disclosure before firms fully anticipate the GPG Mandate. Additionally, we leverage the difference between financial and ESG disclosure to “revisit important empirical relations in the context of CSR reporting,” as suggested in [Christensen et al. \(2021b\)](#). Though we focus on gender diversity, we believe our findings could be generalizable to other ESG issues.

Second, our main result has an important implication for ESG ratings, which purport to help users evaluate firm ESG performance and which take voluntary disclosure as a key input. Recent papers find high disagreement among ESG ratings ([Chatterji et al., 2016](#); [Berg et al., 2020](#); [Serafeim and Yoon, 2021](#)) and find that more ESG disclosure leads to higher disagreement ([Christensen et al., 2021a](#)). We add to this discussion on the connection between ESG disclosure and ratings by showing that when firms with better gender pay gaps are less likely to provide voluntary gender diversity disclosure, these firms receive worse social scores. Our results advise caution when relying on voluntary ESG disclosure to evaluate a firm’s ESG performance.

Finally, we are among the earlier papers to study the setting of the 2017 UK GPG disclosure mandate. Other concurrent papers examine the real effects of this mandate on wages and the gender pay gap (e.g., [Blundell, 2020](#); [Raghunandan and Rajgopal, 2021](#)). [Raghunandan \(2022\)](#) studies misreporting in the gender pay gap data and finds that most misreporting happens among smaller firms. By contrast, we focus on the FTSE 350, where data quality is relatively high, and we use reported GPG data to proxy for a firm’s latent gender diversity performance when it provided voluntary disclosure before the mandate. Our finding demonstrates some limitations of voluntary disclosure (e.g., selective disclosure) and highlights the importance of mandatory disclosure to facilitate better comparison of firms’ gender diversity performances.

⁵For example, [Clarkson et al. \(2008\)](#) studies environmental disclosure and performance based on toxic waste amounts. In their setting, firms provide disclosures knowing their toxic waste amounts will be revealed with a two year lag.

2 Institutional background and research design

2.1 Institutional background

In the UK, public figures and investors began emphasizing workforce gender diversity more strongly in the early 2010s. In 2010, Dame Helena Morrissey founded the 30% Club to call for the boards of the FTSE 100 to be composed of at least 30% women ([30PercentClub, 2020](#)).⁶ In 2011, Lord Mervyn Davies published the Davies Review, urging the FTSE 100, and later the FTSE 350, to improve the proportion of women on their boards to 25%. In 2013, the UK began requiring all quoted firms to report the gender compositions of their boards, senior managers, and total employees on an annual basis.⁷ Amid this backdrop, firms were choosing whether and how to provide voluntary disclosure on gender diversity. If investors and the media believed a firm had poor gender diversity, then a firm risked facing shareholder activism and negative media attention ([BBC, 2011](#)). In July 2015, Prime Minister David Cameron announced the government was considering a gender pay gap disclosure mandate for large firms, but firms did not know the exact metrics to be reported, how the reports would be disclosed, or when reports would be required ([Mason and Treanor, 2015](#)). Then in February 2016, the government announced GPG disclosure would be delayed until 2018 ([Mason, 2016](#)).

In April 2017, the UK passed GPG Mandate, or “Equality Act 2010 (Gender Pay Gap Information) Regulations 2017.” The mandate requires firms to provide annual disclosure about their gender pay gaps through seven specific metrics: (1) the mean gender pay gap, (2) the median gender pay gap, (3) the mean bonus gender pay gap, (4) the median bonus gender pay gap, (5) the proportion of men who are paid a bonus, (6) the proportion of women who are paid a bonus, and (7) the proportion of men and women in each quartile pay band. GPG Mandate applies to UK firms with at least 250 employees as of April that year. Firms were required to submit their first year of GPG data as of April 2017 by April 2018. Firms submit this data to the government and must publish the data on their websites, along with a written statement signed by a director or equivalent-level employee.

⁶The FTSE 100 is made up of the 100 largest firms by market capitalization whose primary listing is the London Stock Exchange.

⁷The Companies Act 2006 (Strategic Report and Directors’ Report) Regulations 2013, SI 2013/1970.

2.2 Research design

We examine whether firms with better GPGs are more or less likely to provide voluntary gender diversity disclosure in annual reports. Figure 1 illustrates the Libby Box framework underlying our research design. The primary relation we examine is between gender diversity performance and the likelihood of voluntary gender diversity disclosure. To operationalize gender diversity performance, we use a firm's gender pay gap reported for April 2017; we assume this GPG is similar to the GPG in 2016 when the firm releases its annual report. To operationalize the likelihood of voluntary disclosure, we use whether a firm provides any voluntary disclosure on gender diversity in the annual report released in 2016.

Our research design relies on two key assumptions. First, we assume a firm's gender pay gap is an important component of its gender diversity, allowing us to use GPG as a measure of a firm's latent gender diversity performance. We believe this assumption is reasonable because GPG is affected by both gender composition and pay by gender at each job level. The UK Government Equalities Office ("GEO") also views GPG as closely linked to gender diversity. For example when giving guidance to firms on GPG Mandate, the GEO states, "Used to its full potential, gender pay gap reporting is a valuable tool for assessing levels of equality in the workplace, male and female participation, and how effectively talent is being maximized" (GEO 2019, p. 4).⁸

Despite the institutional importance of GPG, there are other important components of gender diversity. For example, though not the primary target of GPG Mandate, the regulation also requires firms to report bonus data and gender composition across pay quartiles. Relative to GPG, which implicitly includes bonus pay, bonus pay metrics are more difficult to compare across firms because the ratio of bonus pay to overall pay varies by business model and sector. In addition, we focus on GPG rather than composition across pay quartiles because while it is tempting to think of each pay quartile as corresponding to a broad job level, we do not have enough information to map pay quartile composition to job-level composition. For example, we cannot observe if a male senior manager is paid more and included in the top pay quartile while a female senior manager

⁸This report can be found at:
https://globalhealth5050.org/wp-content/uploads/2019/03/Managing_gender_pay_reporting_07.02.191.pdf

is paid less and included in the upper-middle pay quartile. This difficulty in interpreting the pay quartile composition data also hinders comparability. As a result, though there are other inputs to gender diversity, we choose to focus on GPG because it is institutionally important and the most comparable across firms.

Second, we assume a firm's reported GPG is a good proxy for its GPG one year earlier when it released its annual report in 2016. One potential concern is that between 2016 and 2017, firm GPGs changed, potentially because firms anticipated GPG Mandate. Though we lack reliable GPG data before 2017, we can examine whether firm gender composition changed markedly over time; if it did not, then it is likely GPG also did not change over time. In Figure A1, we split firms into terciles within-industry based on the proportion of female senior managers and employees in 2015, and these terciles change very little over the next two years. In addition, prior literature shows that GPG is shaped by industry-specific labor supply conditions and management's tone at the top, which are both slow-moving.⁹ For example in 2017, Persimmon Group cites the historically male-dominated construction industry as a reason why "opportunities to increase female representation at the most senior grades occur less frequently... so it is very difficult to make a quick impact on the gender pay gap." Taken together, these reasons help alleviate the concern that a firm's GPG differs significantly between 2017 and when the firm released the annual report in 2016.

3 Data and descriptive statistics

3.1 Sample

Our sample covers the FTSE 350, the 350 largest firms by market capitalization whose primary listing is on the London Stock Exchange as of the end of 2016. We begin with this sample for three reasons. First, these firms are more likely to have sophisticated investor bases that can coordinate efforts to push for gender diversity improvements. These firms are thus more likely to use voluntary disclosure to address gender diversity. Second, these firms are larger and more likely to be treated

⁹The causes of the gender pay gap have been widely studied in the labor economics literature. Decomposing the contributing factors to the gender pay gap using US data in 2010, [Blau and Kahn \(2017\)](#) find that work experience explains 14% of the gap, industry and occupation explain 50%, and 38% of the variation remain unexplained. Using data from Uber drivers, [Cook et al. \(2021\)](#) find that the gender pay gap can be explained by differences in work experience and risk aversion by gender. Overall these factors relate to variations in the labor supply.

by GPG Mandate. Third, these firms are likely to have higher media visibility and are more likely to publish high quality gender pay gap data.¹⁰

Table 1 summarizes the rest of our sample selection. We collect all firms' gender pay gap reports from company websites where applicable and drop 123 firms not treated in the first year of GPG Mandate. Of these 123, most are not treated because neither the parent firm nor any subsidiaries have more than 250 employees in the UK. This situation often occurs when the firm is listed on the London Stock Exchange but operates mainly outside the UK. Examples include NMC Health PLC, which operates in the United Arab Emirates, and Hikma Pharmaceuticals PLC, which operates in Jordan. There are 227 firms remaining for which we would like to collect fiscal 2015 annual reports released in 2016. We exclude 7 firms that do not release an annual report in 2016; this occurs when a firm goes public in early 2016, for example. Of the remaining 220, 7 firms either voluntarily disclose gender pay gap data or mention the prospect of mandatory GPG reporting in their annual report. We also exclude these firms because their voluntary disclosure is more likely to reflect anticipation.¹¹ There are 213 firms remaining for which we can analyze voluntary disclosures.

Because annual reports are at the consolidated firm level, we lose some observations that lack consolidated GPG data. A nuance of GPG Mandate is that the 250 employee threshold applies at the legal hiring entity level, but not at the consolidated firm level. For example, Unilever PLC has two main hiring entities in the UK: Unilever UK Ltd and Unilever UK Central Resources Ltd. Each subsidiary is treated by GPG Mandate because it has more than 250 employees, but Unilever PLC, the parent, is technically not treated. Thus by law, Unilever PLC only needs to report GPG data for Unilever UK Ltd and Unilever UK Central Resources Ltd. In Unilever's case, the parent voluntarily reports full GPG data for the consolidated entity. This nuance in the legislation limits our eventual sample because we can only use GPG when it applies to the consolidated firm.¹² Consequently for our descriptive results, we begin with 126 firms, as shown in Table 2 Panel A. We drop industries

¹⁰Raghunandan (2022) studies misreporting in median GPG data and finds that misreporting is mostly concentrated among smaller firms. We check our sample and identify three firms with potentially misreported median GPG. Our results remain robust to excluding these three firms.

¹¹In untabulated analysis, the results are quantitatively similar when we include these 7 firms.

¹²In untabulated analysis, we find that the choice to provide consolidated GPG data within-industry is not correlated with gender composition and is only weakly positively correlated with size, though the coefficient is not statistically significant. This helps address the concern of selection bias.

with fewer than 3 observations, resulting in 123 firms shown in Table 2 Panel B.¹³ Finally in our empirical tests, we also exclude firms without the necessary data for our control variables, leaving us with 116 firms in the sample. Given the reduction in sample size due to data limitations, a possible concern is that our sample may not be representative of the full FTSE 350. In Appendix Table A2, we compare voluntary disclosures of the 123 firms used in our descriptive statistics to the broader set of FTSE 350 firms with annual reports. Overall, we find voluntary disclosure behavior is similar across the two groups.

3.2 Gender pay gap data

We have two sources for gender pay gap data. The first is the official UK government dataset from the Office for National Statistics (ONS).¹⁴ This dataset contains the GPG metrics submitted by treated firms to the UK government. As previously discussed, GPG Mandate technically treats only hiring entities with more than 250 employees. As described in the Unilever example, a parent with fewer than 250 direct employees but with more than 250 cumulative employees across subsidiaries would technically not be treated. The ONS dataset only includes GPG data for entities that are technically treated. The second source is a firm's gender pay gap report containing 2017 data, which we download from the company website. A typical GPG report contains the required metrics, a narrative explaining the metrics, and sign-off from a company representative. Most firms that are treated in any fashion (e.g., an untreated parent entity with treated subsidiaries) voluntarily provide consolidated GPG metrics in their reports. We think this data is credible because senior executives sign-off on the accuracy of the reports. For firms large enough to be included in the FTSE 350, this sign-off carries significant reputation risk if the data turns out to be inaccurate. We use the consolidated data from company GPG reports as our source for GPG wherever possible. The remaining firms that are treated in some capacity but do not voluntarily disclose consolidated data fall into one of three groups: (1) the parent entity is treated and the parent has no subsidiaries; (2) the parent entity is not treated but it has only one subsidiary, which is treated; and (3) others. For group (1), we use the ONS GPG data for the treated entity. This data should work well for

¹³In untabulated analysis, we drop industries with fewer than 5 observations and the results remain similar.

¹⁴This data can be found here:

<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/annualsurveyofhoursandearningsashegenderpaygaptables>

group (1) where by definition, the firm has only one entity. However there is a slight concern that the ONS GPG data does not accurately reflect group (2). By definition, this group has less than 250 employees at the parent, but those employees are likely to be in more administrative functions and thus paid differently than employees at the subsidiary. For this reason we do not include group (2) in the main analysis but include the group in untabulated robustness tests. We do not have an estimate for consolidated GPG for group (3), so we exclude these firms.

Across both sources, GPG is calculated as:

$$\frac{(A - B)}{A}$$

A is the mean hourly rate of pay of all male full-pay relevant employees and B is the mean hourly rate of all female full-pay relevant employees. We construct $Good\ GPG = 1 - GPG$ so a narrower GPG results in a higher $Good\ GPG$.¹⁵

Table 2 Panel A summarizes $Good\ GPG$ across 2-digit NAICS industries in our sample. The order of industries in the table is ranked from the best to the worst average $Good\ GPG$. Consistent with expectations, industries like finance and mining have relatively high gender pay gaps (low $Good\ GPG$). On average, the mean hourly rate of pay for female workers in the finance and insurance industry is 32% less than that of male workers. Overall, the average GPG is 20% among our sample of 126 firms. In the first column, we provide the average $Good\ GPG$ based on all entities in the ONS dataset. The average GPGs between our sample and the ONS sample are similar except for in the mining and healthcare industries where we have fewer than 3 observations. In subsequent analysis, we drop these two industries, resulting in 123 firms.

3.3 Voluntary gender diversity disclosure in annual reports

We measure voluntary gender diversity disclosure from annual reports released in 2016. Recall the UK passed regulation in 2013 mandating the disclosure of gender composition data at the board, senior manager, and total employee levels. In addition, the Davies Review called for firms to describe

¹⁵We use the signed version of GPG because a higher pay for female is generally considered favorable. Our results remain robust when we use the absolute value of GPG in Appendix A4.

their board nomination process. Consequently, annual reports for quoted UK firms typically contain both these required disclosures and additional commentary around board composition (Davies, 2021). We consider disclosures of this flavor to be mandatory and purposefully exclude them from our set of voluntary gender diversity disclosures. For example, we consider “gender diversity” in a paragraph on the board nomination process to be mandatory. This quirk of the setting makes it more difficult to apply textual analysis to categorize the disclosures in the reports. An algorithm would need to distinguish between mandatory and voluntary disclosures that often use the same word or phrase. Thus we resort to hand-collection. We provide a detailed description of the data collection process in Appendix B.

We find substantial variation in what firms voluntarily disclose about gender diversity in their annual reports. Table 2 Panel B presents the summary statistics of these disclosure characteristics. Among the 46% of firms that provide voluntary disclosure, the average firm discloses 125 words. Using VADER (Valence Aware Dictionary and Sentiment Reasoner) sentiment analysis in python, we see the words disclosed are mostly neutral, with a neutral sentiment score of 0.842, relative to a positive sentiment score of 0.149 and a negative sentiment score of 0.008.

Based on UK GEO guidelines and how firms commonly disclose about gender diversity, we categorize disclosure into five forms: (i) metrics, (ii) targets, (iii) external commitments, (iv) internal actions, and (v) policy benefits.¹⁶ Metrics includes gender diversity metrics such as the percentage of women in each division or the percentage of women earning above a certain salary. Metrics do not include the mandatory gender composition metrics. Targets include quantitative gender diversity targets with a target year. External commitments include commitments to external gender diversity organizations. Internal actions include internal actions the firm is taking to improve gender diversity. Finally, policy benefits include benefits employees receive that firms explicitly link to encouraging gender diversity. We code the categories as indicator variables. If a firm uses a certain form of voluntary disclosure, we code that category as 1; otherwise we code that category as 0.

¹⁶In guidelines for implementing gender pay reporting, the Government Equalities Office suggests steps firms can take to improve their gender pay gaps. These suggestions include “talent management and people development” to ensure the “pipeline is working for senior women” and “flexible working” (GEO 2019, p. 31). Internal actions and policy benefits mirror these suggestions, respectively. Outside of the government, the 30% Club recommends firms have a “measurable goal with a defined timetable” and a “concerted and consistent series of actions and programmes” (30PercentClub, 2020). Metrics, targets, internal actions, and policy benefits reflect these recommendations.

Table 2 Panel B shows the most common categories are internal actions at 33% of the sample, metrics at 24%, and policy benefits at 13%. The least common categories are targets at 7% and external commitments at 7%. Within a category, there is also considerable variation. The following examples demonstrate the range of information firms disclose in each category.

First, voluntarily disclosed metrics run the gamut from measuring women in leadership and women entering the firm to women earning a certain salary. For example in its annual report released in 2016, SSE PLC, an energy firm, discloses:

- *Of the available workforce in the UK and Ireland 51% are women, yet only 30% of the people employed in SSE are women. Of the senior management team, 30% are women. There are 66 direct reports to SSE's Executive Directors and Managing Directors – 20 women and 46 men.*
- *Only 10% of SSE's current female workforce are in positions that pay £40,000 FTE (full time equivalent) or more.*
- *SSE has a wide variety of flexible working arrangements in place across all of its businesses. This includes over 1900 employees (85% women: 15% men) working some form of part-time hours to meet their personal circumstances.*

However voluntarily disclosed metrics can also include metrics on women entering internal development programs. For example GlaxoSmithKline PLC, a pharmaceuticals firm, discloses:

- *As at 31 December 2015, women represented 52% of recruits to our Future Leaders programme.*
- *In 2015, 118 more female managers began our Accelerating Difference programme for high performing women leaders.*

Second, voluntarily disclosed targets can be applied to women at both entry and senior levels and at certain pay bands. For example SSE PLC discloses:

- *Only 10% of SSE's current female workforce are in positions that pay £40,000 FTE (full time equivalent) or more. SSE has set an overall target to increase this to 25%*

by 2025.

Third, voluntarily disclosed external commitments contain the least within-category variation where firms disclose joining or signing onto UK gender diversity organizations such as the 30% Club or Women on Boards. For example Pearson PLC, an education publishing firm, discloses:

- *We remain an enthusiastic member of the 30% Club which brings together chairs and CFOs to work together on gender balance.*

Fourth, internal actions often include unconscious bias training and programs to encourage women through the corporate pipeline. For example Pearson PLC discloses:

- *Highlights of our activities include: Raising awareness about the impact of unconscious bias on key people management decisions. To date, over 4,000 employees have completed our interactive training on the topic.*
- *We participate in their cross-company mentoring programme which helps the development of talented mid-career women.*

However internal actions also include promoting networking events for women, engaging in outreach to female students, and changing recruiting policies. For example Babcock International Group PLC, an engineering firm, discloses:

- *In considering the issues around gender diversity we can report that Babcock has... focused our graduate recruitment programme, particularly of engineering graduates, on those universities that have a richer undergraduate gender mix so as to improve the diversity of the pool of talent from which we recruit our engineers and managers of the future - female graduates now make up 22% of the annual intake, an increase of 2% on last year.*

Fifth, disclosing policy benefits includes talking about flexible working or parental leave options and connecting these employment benefits to gender diversity. For example regarding flexible working, Babcock International Group also discloses:

- *In considering the issues around gender diversity we can report that Babcock has...*

continued to actively consider how to make management roles more attractive and amenable to females so as to increase the numbers of women in management. A review of flexible working practices in other companies is being undertaken to consider what is likely to have the most positive impact.

Additionally touching on parental leave, OneSavings Bank PLC discloses:

- *We also carried out a review to identify if there were any impediments to ensuring gender balance across the business and looked at recruitment, retention and promotions policy and procedure. Although no specific barriers were identified, maternity and paternity benefits have been improved and training in relation to unconscious bias is to be introduced across all management grades, as a result of this review.*
- *The Bank will be tracking the take up of flexible working arrangements and other gender related metrics going forwards, and will assess whether any further measures or measurable objectives need to be put in place.*

4 Main results

4.1 GPG and voluntary disclosure

To descriptively examine which firms provide voluntary gender diversity disclosure, we separate firms into terciles within the same industry based on *Good GPG*.¹⁷ On average, *Worse GPG*, *Medium GPG*, and *Better GPG* firms have a *Good GPG* of 0.71, 0.83, and 0.90, which translates to a gender pay gap of 29%, 17%, and 10%, respectively. Table 2 Panel B shows the descriptive statistics of disclosure variables split by *Good GPG* terciles, Table 2 Panel C shows the correlation matrix of the disclosure variables and *Good GPG*, and Figure 2 plots the average disclosure rates by *Good GPG* terciles. Together, these reveal three patterns. First, the rate of voluntary disclosure is low: 46% of firms provide voluntary disclosure but 54% do not. Second, *Worse GPG* firms disclose more often than *Medium GPG* and *Better GPG* firms, at 63% of the time compared to 36% and 34%. Third, disclosure behavior varies across the disclosure categories.

¹⁷Given our limited sample size, we use terciles to enhance power and observe any nonlinear patterns.

If disclosure is costless and truthful, then traditional disclosure models predict that all firms will disclose because firms with better performance will disclose to separate themselves from others, resulting in full unraveling (Milgrom, 1981; Grossman, 1981). Our first pattern of incomplete disclosure runs counter to unraveling, suggesting disclosure is costly. This pattern also appears in other settings; for example, Clarkson et al. (2008) find a 64% rate of voluntary environmental disclosure. If ESG disclosure is similar to financial disclosure, we might expect to find that firms with better performance are more likely to disclose, as predicted by models that incorporate costly disclosure (Verrecchia, 1983). However, our second pattern suggests firms with better GPGs are actually less likely to provide voluntary disclosure.

In a formal regression, we estimate the relationship between *Good GPG* and *Disclosure*, which is one of three disclosure outcomes. We use the following empirical specification:

$$Disclosure = \beta_0 + \beta_1 Good\ GPG + \beta_2 Controls + \beta_3 Industry\ FE + \epsilon$$

Our variable of interest is β_1 , which captures how likely firms with higher *Good GPG* are to provide *Disclosure*. If providing voluntary disclosure implies having a higher *Good GPG*, then β_1 will be positive.

We include industry fixed effects by 2-digit NAICS sector code because it is likely that an industry's labor supply conditions contribute to a firm's gender pay gap and strategic voluntary disclosure choice. For example, construction firms may have larger gender pay gaps than retail firms because there are relatively fewer female employees entering the construction industry. We control for firm size using the log market capitalization from Worldscope because larger firms may face more media scrutiny and be more likely to disclose. We include the percentage of female directors from Asset4 to control for existing mandatory disclosures on composition in case firms choose their voluntary disclosures with mandatory disclosures in mind. We include the environmental score from Asset4 to proxy for the extent of firms' environmental disclosures, capturing drivers of voluntary non-financial disclosure unrelated to gender diversity. The control variables are measured in 2015.

Table 3 Panel A presents our results for the three main disclosure outcomes: an indicator for any voluntary disclosure called *GD Disclosure*, the number of categories disclosed, and the number

of words used in the voluntary disclosure. Column 1 shows that the coefficient on *Good GPG* is negative and statistically significant, which suggests that firms with better GPGs are less likely to provide voluntary disclosure. Specifically, having a 10% better GPG is associated with a 17% lower likelihood of disclosure (-1.7×0.1). As an alternative interpretation, a one standard deviation increase in *Good GPG* (11.6%) is associated with a 20% lower likelihood of disclosure (-1.7×0.116). Column 2 shows that firms with better gender pay gaps also disclose through fewer categories. While the magnitude in column 3 suggests these firms also use fewer words related to gender diversity, the coefficient is not statistically significant.

We conduct a battery of robustness tests on our main results. In Appendix Table A4, the results remain similar when we take the absolute value of mean GPG when calculating *Good GPG*. In Appendix Table A5, the results remain similar when we use 3-digit NAICS instead of 2-digit NAICS as the industry classification. This tighter industry classification provides some comfort that we are not picking up industry variations in GPG. However, this test reduces the sample size further as there are fewer industry groups with multiple firms. In Appendix Table A6, we find qualitatively similar results when we use the median GPG to calculate *Good GPG*: the coefficients remain negative but are not statistically significant. One reason for this is that the median GPG captures less variation in the range of earnings, while the mean GPG includes the impact of very low or very high-paying positions.

The third pattern from Table 2 Panel B shows *Worse GPG* firms disclose metrics and policy benefits most often but *Medium GPG* firms disclose targets most often. In Table 3 Panel B, we repeat the analysis in Panel A for each disclosure category. While in all specifications the coefficient on *Good GPG* is negative, it is only statistically significant for metrics. We find that external commitments, internal actions, and benefits are not statistically significantly associated with *Good GPG*. These three disclosure categories are relatively qualitative. In contrast, the more quantitative category of metrics does seem to relate to *Good GPG*. Specifically, column 1 shows that having a 10% better GPG translates to a 9% lower likelihood of disclosing metrics (-0.859×0.1). While the coefficient is not statistically significant for targets, one reason may be that target disclosure exhibits non-linearity: the descriptive statistics show *Medium GPG* firms are most likely to disclose

targets. We examine metrics in more detail as part of potential explanations.

4.2 Implications for ESG ratings

An important implication from our findings is that firms with better ESG performance may not be the firms that provide more voluntary ESG disclosure. Many investors rely on ESG ratings to evaluate a firm's ESG performance (Hartzmark and Sussman, 2019); however, recent work finds low correlation across ESG ratings from different rating providers (Chatterji et al., 2016; Berg et al., 2020). One source of variation comes from providers placing different weights on firm disclosure. For example, some ratings penalize firms for not providing ESG disclosure on a particular topic, and firms with more disclosure tend to receive higher ESG scores (Khan et al., 2016; Li et al., 2021). Given our finding that firms with better GPGs are less likely to provide voluntary disclosure, we hypothesize that ratings more reliant on voluntary disclosure are more negatively associated with *Good GPG*.

To test this hypothesis, we use three popular ESG ratings that use voluntary disclosure as an input to different extents: Asset4 (from Thomson Reuters), Sustainalytics (from Morningstar), and MSCI. From Asset4, we use the social score measured at the end of fiscal 2015. The social score is a broad rating for a firm's social performance, including its workforce gender diversity. From Sustainalytics, which provides more frequent monthly data, we use the social score measured in June 2016. From MSCI, which also provides monthly data, we use the social score measured in June 2016. During our time period, both Sustainalytics and MSCI used a combination of firm disclosure and controversy assessments based on news media while Asset4 scores were predominately based on voluntary disclosure (ThomsonReuters, 2013; Sustainalytics, 2017; MSCI, 2017). For example in Asset4's social score methodology, when a key performance indicator ("KPI") is missing a response, the KPI is assigned either a value of 0 or the minimum value in that firm's industry or region (ThomsonReuters 2013, p. 11).¹⁸

In Table 4 Panel A, we repeat our main specification from Table 3 but replace the outcome variable with the ESG ratings. To enhance comparability, we only use observations with ESG ratings

¹⁸This methodology discussion can be found at: <https://www.thomsonreuters.com/content/dam/openweb/documents/pdf/tr-com-financial/methodology/corporate-responsibility-ratings.pdf>

from all three providers. When controlling for environmental score, we use the environmental score of the respective ESG rating provider. Among the three ESG rating providers, the coefficient on *Good GPG* is negative and statistically significant for the social score from Asset4 in column 1. The result in column 1 suggests that having a 10% better GPG is associated with a 3% lower social score (0.214/10/0.619). Because Asset4's scores rely heavily on voluntary disclosure as inputs, this result is consistent with our main finding of a negative relation between *Good GPG* and voluntary disclosure. In contrast, the coefficients of interest for Sustainalytics and MSCI in columns 2 and 3 are not statistically distinguishable from zero, suggesting the rating is not associated with a firm's gender pay gap.

One potential reason the negative correlation between ESG rating and *Good GPG* is only statistically significant for Asset4 is Asset4's higher reliance on voluntary disclosure. Because ESG rating providers do not provide exact details on the weights they place on voluntary ESG disclosure, we regress the ESG ratings on our voluntary disclosure categories. Consistent with our expectations, in Table 4 Panel B, the disclosure categories are significantly positively correlated with the Asset4 social score. A few categories are also significantly positively correlated with the Sustainalytics social score, but two categories are also negatively correlated. If we compare the adjusted within R-squared, which represents how much the disclosure categories explain the variation in ESG ratings, this statistic is the highest for Asset4 ratings. Overall, our findings join recent work in advising caution when using ESG ratings to assess ESG performance.

5 Potential explanations

In this section, we attempt to understand why there is a negative relation between voluntary disclosure and *Good GPG*, which runs opposite to predictions from traditional financial disclosure theories. We explore potential reasons based on the differences between ESG and financial disclosure (Christensen et al., 2021b).

One difference is that relative to financial information, ESG information is relevant to a broader set of stakeholders who can learn about ESG performance from sources outside of financial disclosure. In our setting, employees have historically used labor market information such as workplace rankings

and job review websites to learn about a firm's gender diversity performance. If firms with better GPGs have already been recognized by these labor market information sources, then they may not need to provide their own voluntary disclosure. To study this possibility, we proxy for labor market information using *The Times* Top 50 Employers for Women award. Since 2006, *The Times* has annually recognized 50 firms in the UK that are considered the best employers for women. To be considered for the award, firms respond to a survey conducted by the award organizer. The survey asks for descriptive details on the firm's gender diversity efforts, including questions such as, "What are you doing to recruit gender diverse workforce?" and "How do you prevent and tackle sexual harassment and inappropriate behavior?" Responses are reviewed by experts who select the top 50 employers for women.

In Table 5 Panel A, we estimate our main empirical model and replace the disclosure variable with an indicator for receiving *The Times* Top 50 Employers for Women award by 2017. This label applies to 15 firms in our sample. To enhance power in studying potential explanations, we use the terciles of *Good GPG* as the dependent variable. Column 1 shows both *Medium GPG* and *Better GPG* firms are statistically significantly more likely to receive this award than *Worse GPG* firms, and the F-test shows there is no significant difference between the coefficients for *Medium GPG* and *Better GPG*. This result suggests that the availability of labor market information may explain why firms with better GPGs are less likely to provide voluntary disclosure.¹⁹

Another difference between financial and ESG disclosure is in the credibility of voluntary disclosure. Unlike financial reporting, where there are standard accounting measures, ESG reporting involves a diverse set of measurements. Additionally, most ESG disclosures are voluntary and are not verified by later mandatory disclosure (Ball et al., 2012).²⁰ In a disclosure environment with multiple potential measures and with few disciplining mechanisms, firms with worse GPG may engage in selective disclosure.

We investigate this possibility by examining metrics in detail because metrics drive the overall finding of higher disclosure among firms with worse GPG. The metric most commonly disclosed

¹⁹We acknowledge this interpretation is hindered by the small sample size of firms receiving the award. We are in the process of collecting other labor market information relating to gender diversity.

²⁰GPG Mandate is one of the few mandatory non-financial disclosures with well-specified metrics.

is the percentage of women in leadership positions. In Table 5 Panel B column 1, we replace the dependent variable in the main specification with this metric; we find a similar pattern as in our main results where firms in the *Worse GPG* group are more likely to voluntarily disclose. In column 2, we interact each tercile with the percentage of females in the top pay quartile for each *GPG* group; this data comes from GPG Mandate and hence is available even for firms without voluntary disclosure. The results show that in the *Worse GPG* group only, firms with relatively more women in the top pay quartile are more likely to voluntarily disclose metrics related to women in leadership. This result suggests firms in the *Worse GPG* group with relatively *more* women in leadership may be selectively disclosing metrics that appear more favorable to them.

We caution against over-interpreting these potential explanations because our sample size is small. Additionally, we acknowledge there may be other reasons why firms with better GPGs are less likely to provide voluntary disclosure. In Appendix Tables A7 and A8, we explore two other potential explanations. First, we explore if prior to GPG 2017, investors could identify firms with worse GPGs and pressure them to disclose more on gender diversity. From news articles in 2014 and 2015, we observe that when investors pressured firms about gender diversity, they focused on poor board gender composition because at the time, the FTSE 350 faced calls from public figures to reach 25% women on their boards (Davies, 2021). If investor pressure explains our result, then we would expect firms with less than 25% women on their boards to be more likely to provide voluntary disclosure. In Appendix A7, we do not find evidence that they do. Second, firms with better GPGs may be less likely to disclose because these firms have less room for improvement. This explanation applies more specifically to disclosure of gender diversity targets because a firm may not need to set a target if it is already performing relatively well. In Appendix A8, we study this possibility by examining the eight firms that voluntarily disclose targets. We find five of these firms have targets that reflect room for improvement, and interestingly, none of these firms have better GPGs relative to their industry peers. Two firms disclose targets and have better GPGs relative to peers, but these targets do not reflect room for improvement because they are unrelated to workforce diversity; for example, Unilever's target is to empower five million women by 2020, which is unrelated to its workforce. These targets suggest room for improvement may be a plausible

explanation, but interpretation is difficult because of the small sample size and low frequency of target disclosure.

6 Conclusion

We examine if firms with better ESG performance are more or less likely to provide voluntary ESG disclosure, and we study this research question in the context of the gender pay gap. We use a firm's gender pay gap revealed after a 2017 UK mandate to proxy for a firm's gender diversity performance when it releases its annual report in 2016.

We find that firms with better GPGs are less likely to provide voluntary disclosure about gender diversity. Our results provide practical insights on how to evaluate a firm's gender diversity performance. Our findings suggest that the choice of providing voluntary ESG disclosure is not always correlated with better ESG performance. We find that firms with better gender pay gaps are more likely to have worse Asset4 social scores, which rely heavily on voluntary disclosure. We find essentially no relationship between a firm's Sustainalytics and MSCI social scores and gender pay gap. Thus the reliance of ESG ratings on voluntary disclosure can weaken the informativeness of the ratings themselves when there is a negative relationship between performance and disclosure.

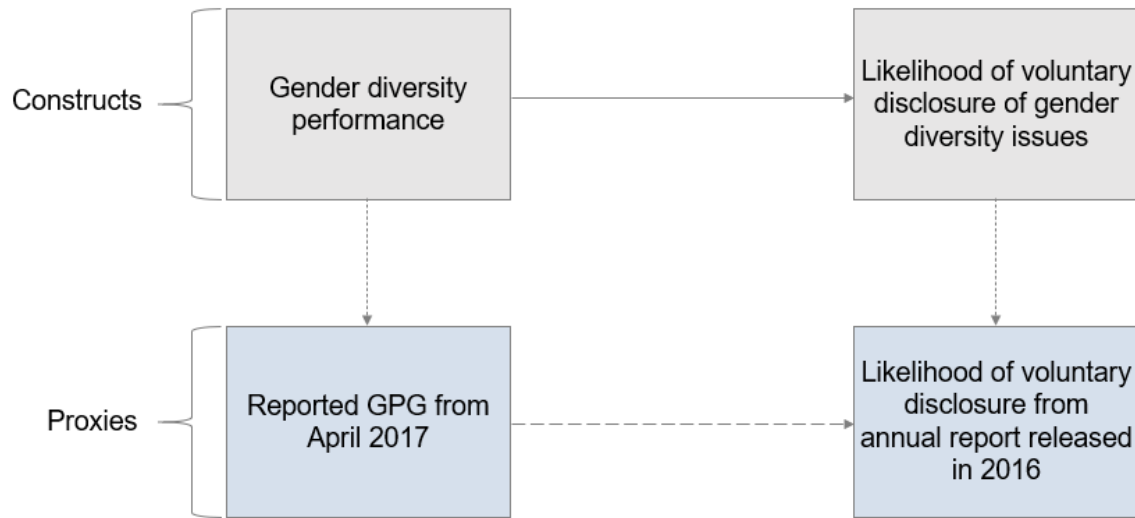
Finally, we provide potential explanations for this negative relation by considering the difference between ESG and financial disclosure. We show that some of these differences, such as the diverse user base of stakeholders and diverse measurements for ESG information, provide different predictions for ESG disclosure than for traditional financial disclosure. Because ESG disclosure can be less credible and stakeholders can learn about gender diversity from sources such as *The Times* Top 50 Employers for Women list, firms with better GPGs may not need to disclose.

References

- 30PercentClub (2020). United Kingdom Chapter. *30 Percent Club*.
- Al-Tuwaijri, S. A., Christensen, T. E., and Hughes II, K. (2004). The relations among environmental disclosure, environmental performance, and economic performance: a simultaneous equations approach. *Accounting, organizations and society*, 29(5-6):447–471.
- Ball, R., Jayaraman, S., and Shivakumar, L. (2012). Audited financial reporting and voluntary disclosure as complements: A test of the confirmation hypothesis. *Journal of accounting and economics*, 53(1-2):136–166.
- BBC (2011). Firms ‘ignore Davies report targets for women’. *BBC News*.
- Berg, F., Koelbel, J. F., and Rigobon, R. (2020). Aggregate confusion: The divergence of ESG ratings. *Available at SSRN 3438533*.
- Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865.
- Blundell, J. (2020). Wage responses to gender pay gap reporting requirements. *Available at SSRN 3584259*.
- Chatterji, A. K., Durand, R., Levine, D. I., and Touboul, S. (2016). Do ratings of firms converge? implications for managers, investors and strategy researchers. *Strategic Management Journal*, 37(8):1597–1614.
- Christensen, D. M., Serafeim, G., and Sikochi, A. (2021a). Why is corporate virtue in the eye of the beholder? The case of ESG ratings. *The Accounting Review*.
- Christensen, H. B., Hail, L., and Leuz, C. (2021b). Mandatory CSR and sustainability reporting: Economic analysis and literature review. *Review of Accounting Studies*.
- Clarkson, P. M., Li, Y., Richardson, G. D., and Vasvari, F. P. (2008). Revisiting the relation between environmental performance and environmental disclosure: An empirical analysis. *Accounting, Organizations and Society*, 33(4-5):303–327.
- Cook, C., Diamond, R., Hall, J. V., List, J. A., and Oyer, P. (2021). The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. *The Review of Economic Studies*, 88(5):2210–2238.
- Crawford, V. P. and Sobel, J. (1982). Strategic Information Transmission. *Econometrica*, 50(6):1431.
- Davies, L. M. (2021). 2011 - 2015 The Davies Review - FTSE Women Leaders.
- Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality. *Journal of Law and Economics*, 24(3):461–483.
- Hartzmark, S. M. and Sussman, A. B. (2019). Do investors value sustainability? a natural experiment examining ranking and fund flows. *The Journal of Finance*, 74(6):2789–2837.
- Ingram, R. W. and Frazier, K. B. (1980). Environmental performance and corporate disclosure. *Journal of Accounting Research*, 18(2):614–622.
- Khan, M., Serafeim, G., and Yoon, A. (2016). Corporate sustainability: First evidence on materiality. *The accounting review*, 91(6):1697–1724.

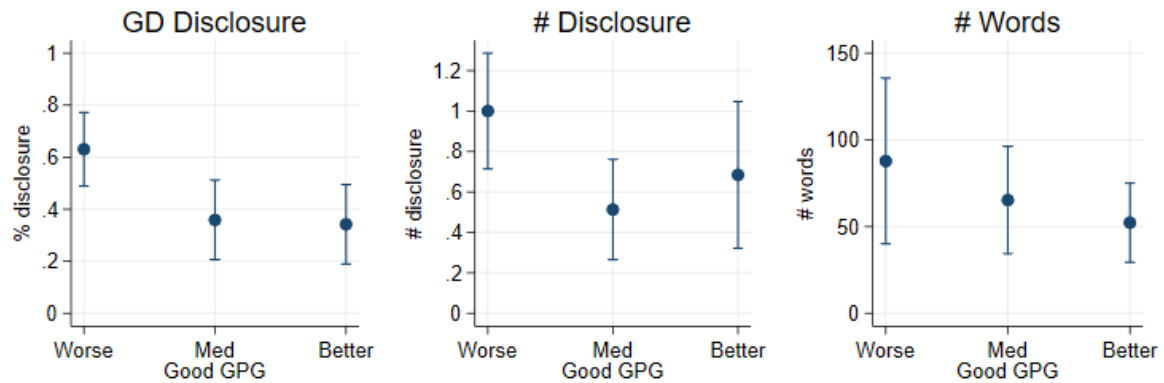
- Kollewe, J. and Butler, S. (2018). Top UK companies under fire over lack of women in boardroom. *The Guardian*.
- Li, J., Lu, S., and Nassar, S. (2021). Corporate social responsibility metrics in S&P 500 firms' 2017 sustainability reports. *Rustandy Center for Social Sector Innovation: The University of Chicago*.
- Marriage, M. (2015). Investors vow to raise diversity on all-male FTSE 250 boards. *The Financial Times*.
- Mason, R. (2016). Gender pay gap reporting for big firms to start in 2018. *The Guardian*.
- Mason, R. and Treanor, J. (2015). David Cameron to force companies to disclose gender pay gaps. *The Guardian*.
- Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, pages 380–391.
- MSCI (2017). Msci esg ratings.
- Patten, D. M. (2002). The relation between environmental performance and environmental disclosure: a research note. *Accounting, Organizations and Society*, 27(8):763–773.
- Raghunandan, A. (2022). Gender pay gap misreporting.
- Raghunandan, A. and Rajgopal, S. (2021). Mandatory gender pay gap disclosure in the UK: Did inequity fall and do these disclosures affect firm value? *Available at SSRN 3865689*.
- Serafeim, G. and Yoon, A. (2021). Stock price reactions to ESG news: The role of ESG ratings and disagreement. *The Review of Accounting Studies (forthcoming)*.
- Sustainalytics (2017). Sustainalytics' ESG rating research methodology.
- ThomsonReuters (2013). Thomson Reuters corporate responsibility ratings: Rating and ranking rules and methodologies.
- Verrecchia, R. E. (1983). Discretionary disclosure. *Journal of Accounting and Economics*, 5:179–194.
- Wiseman, J. (1982). An evaluation of environmental disclosures made in corporate annual reports. *Accounting, Organizations and Society*, 7(1):53–63.

Figure 1: Libby Box framework for research design



This figure illustrates the Libby Box framework underlying our research design. The primary relation we examine is between gender diversity performance and the likelihood of voluntary gender diversity disclosure. To operationalize gender diversity performance, we use a firm's gender pay gap reported for April 2017; we assume this GPG is similar to the GPG in 2016 when the firm releases its annual report. To operationalize the likelihood of voluntary disclosure, we use whether a firm provides any voluntary disclosure on gender diversity in the annual report released in 2016

Figure 2: *Good GPG* and Voluntary Disclosure



This figure plots the relationship between *Good GPG* and voluntary disclosure on gender diversity in annual reports released in 2016. The plots show 95% confidence intervals. $Good\ GPG = 1 - GPG$, where GPG is the mean hourly gender pay gap for the firm; *Good GPG* is grouped in terciles within-industry. *GD Disclosure* is 1 if the firm provides any of the voluntary disclosure categories. *# Disclosure* is the number of disclosure categories. *# Words* is the average number of words related to gender diversity.

Table 1: Sample Selection

	Sample Size
FTSE 350	350
Drop if not treated by GPG Mandate	(123)
Drop if no annual report released in 2016	(7)
Drop firms that mention GPG Mandate in the annual report released in 2016	(7)
Drop firms without GPG data for the consolidated firm	(87)
Final sample	126

This table presents the sample selection. We start with the Financial Times Stock Exchange 350 (“FTSE 350”) at the end of 2016. The FTSE 350 is made up of the 350 largest firms by market capitalization whose primary listing is on the London Stock Exchange. We drop 123 firms not treated in the first year of the UK GPG disclosure mandate (“GPG Mandate”). Of these 123, most are not treated because neither the parent firm nor any subsidiaries have more than 250 employees in the UK. We drop 7 firms that do not release an annual report in 2016 because we use the reports to measure voluntary disclosure (this occurs when a firm goes public in early 2016). We drop another 7 firms that mention the prospect of mandatory gender pay gap (“GPG”) reporting in their annual report because of anticipation concerns. Of the remaining 213 firms, we drop 87 firms without consolidated GPG data. A nuance of GPG Mandate is that the 250 employee threshold applies at the legal hiring entity level, but not at the consolidated firm level. For example, Unilever PLC has fewer than 250 employees, so it is not treated, but its two main subsidiaries in the UK have more than 250 employees and are treated. By law, Unilever PLC only needs to report GPG data for its subsidiaries, but Unilever PLC voluntarily reports GPG data for the consolidated entity. This nuance in the legislation limits our sample because we can only use GPG when it applies to the consolidated firm. This leaves us with 126 firms, as shown in Table 2 Panel A. Later in Table 2 Panel B, we drop industries with fewer than 3 observations (affects 3 firms), resulting in 123 firms. Finally in our empirical tests, we exclude 7 firms without control variable data, resulting in 116 firms.

Table 2: Descriptive Statistics

<i>Panel A: Good GPG by Industry</i>							
	UK		N	Our Sample			
	Mean	Mean		Std. dev.	P25	P50	P75
Administrative and Support	0.887	0.872	8	0.089	0.792	0.842	0.962
Retail Trade	0.848	0.870	16	0.054	0.826	0.861	0.918
Utilities	0.845	0.864	3	0.082	0.780	0.869	0.944
Arts and Entertainment	0.782	0.849	3	0.014	0.835	0.850	0.862
Manufacturing	0.865	0.842	29	0.097	0.795	0.865	0.909
Scientific	0.818	0.840	8	0.100	0.752	0.813	0.936
Accommodation and Food	0.919	0.832	7	0.115	0.761	0.863	0.919
Transportation and Warehousing	0.895	0.827	4	0.160	0.701	0.813	0.954
Wholesale Trade	0.848	0.816	3	0.101	0.710	0.826	0.911
Information	0.805	0.755	10	0.058	0.722	0.753	0.790
Construction	0.776	0.752	5	0.193	0.656	0.773	0.860
Real Estate	0.834	0.738	4	0.175	0.598	0.750	0.877
Other Services	0.874	0.724	3	0.138	0.565	0.795	0.811
Health Care	0.933	0.711	2	0.076	0.657	0.711	0.765
Finance and Insurance	0.737	0.683	20	0.096	0.602	0.700	0.752
Mining	0.777	0.450	1	.	0.450	0.450	0.450
Total	0.829	0.799	126	0.120	0.729	0.810	0.885

Continued on following page

Table 2 (continued)

<i>Panel B: Disclosure Measures</i>					
	Pooled		Worse GPG	Medium GPG	Better GPG
	Mean	Std. dev.	Mean	Mean	Mean
Good GPG	0.803	0.116	0.707	0.825	0.897
GD disclosure	0.455	0.500	0.630	0.359	0.342
# disclosure categories	1.643	0.841	1.586	1.429	2.000
Word count	124.518	158.014	131.172	129.857	103.923
Positive sentiment	0.149	0.079	0.123	0.126	0.206
Neutral sentiment	0.842	0.081	0.868	0.866	0.785
Negative sentiment	0.008	0.012	0.009	0.007	0.009
Metrics	0.244	0.431	0.370	0.154	0.184
Targets	0.073	0.261	0.043	0.128	0.053
External commitments	0.073	0.261	0.065	0.026	0.132
Internal actions	0.333	0.473	0.326	0.333	0.342
Benefits	0.130	0.338	0.174	0.128	0.079
Top 50 Employer for Women	0.114	0.319	0.022	0.154	0.184
Asset4 social score	0.619	0.181	0.650	0.597	0.606
Sustainalytics social score	0.624	0.105	0.640	0.599	0.635
MSCI social score	0.477	0.164	0.471	0.497	0.464
Observations	123		46	39	38

Continued on following page

Table 2 (continued)*Panel C: Correlation matrix*

	Good GPG	GD disclosure	# disclosure categories	Word count
Good GPG	1.000			
GD disclosure	-0.145	1.000		
# disclosure categories	-0.072	0.809***	1.000	
Word count	-0.029	0.394***	0.614***	1.000

Panel A shows descriptive statistics of *Good GPG* across 2-digit NAICS industry. *Good GPG* = $1 - GPG$, where *GPG* is the mean hourly gender pay gap for the firm. The panel is ordered by average *Good GPG* in each industry (descending). The first column shows the average *Good GPG* for all entities treated by the GPG Mandate. From Panel A to Panel B, we drop industries with fewer than 3 observations. **Panel B** shows descriptive statistics for our disclosure measures and other key variables. In the last three columns, we group firms in terciles of *Good GPG* within-industry. *GD disclosure* is 1 if the firm discloses any of the voluntary disclosure categories. *# Disclosure category* is the number of disclosure categories disclosed by the firm. *Word count* is the number of words related to gender diversity in the firm's disclosure. Sentiment comes from using the python package vaderSentiment. *Metrics* is 1 if the firm discloses a quantitative gender diversity metric. *Targets* is 1 if the firm discloses a gender diversity target with a quantitative goal and time-frame. *External commitments* is 1 if the firm commits to an external gender diversity organization, such as the 30% Club. *Internal actions* is 1 if the firm discloses an internal action targeted towards improving gender diversity, such as gender-blind recruiting. *Benefits* is 1 if the firm discloses diversity-related benefits employees receive, such as offering flexible working. *Top 50 Employer for Women* is 1 if the firm received this award from *The Times* in the UK before 2017. *Asset4 social score*, *Sustainalytics social score*, and *MSCI social score* are social scores from three top ESG rating providers, Asset4 of Thomson Reuters, Sustainalytics of Morningstar, and MSCI. **Panel C** shows the correlations between *Good GPG* and our three main disclosure measures.

*, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 3: Main results

<i>Panel A: Good GPG and Voluntary Disclosure</i>			
	(1)	(2)	(3)
	GD disclosure	# Disclosure categories	Word count
Good GPG	-1.700*** (-3.65)	-2.414* (-1.91)	-220.701 (-1.66)
Size	0.026 (0.63)	0.142 (1.46)	20.441 (1.31)
Female directors	1.213** (2.70)	2.886*** (3.78)	67.802 (1.23)
Environmental score	0.034 (0.19)	0.372 (1.32)	39.872 (0.56)
N	116	116	116
Adj. R-squared	0.141	0.218	0.137
Industry FE	Yes	Yes	Yes
Cluster	Industry	Industry	Industry

Continued on following page

Table 3 (continued)

<i>Panel B: Good GPG and Voluntary Disclosure Categories</i>					
	(1)	(2)	(3)	(4)	(5)
	Metrics	Targets	External commitments	Internal actions	Benefits
Good GPG	-0.859** (-2.77)	-0.105 (-0.35)	-0.051 (-0.18)	-0.383 (-1.00)	-0.646 (-1.54)
Size	-0.036 (-0.73)	0.054*** (3.08)	0.019 (0.50)	0.068 (1.68)	0.048 (1.39)
Female directors	1.092** (2.25)	0.299 (0.86)	0.747 (1.70)	0.708* (1.88)	-0.544 (-1.75)
Environmental score	0.217 (1.02)	0.103* (2.10)	-0.101 (-0.56)	0.624*** (3.06)	-0.107 (-0.56)
N	116	116	116	116	116
Adj. R-squared	0.098	0.039	-0.014	0.178	0.051
Industry FE	Yes	Yes	Yes	Yes	Yes
Cluster	Industry	Industry	Industry	Industry	Industry

This table summarizes our main results. **Panel A** shows the results from estimating the regression $Disclosure = \beta_0 + \beta_1 Good\ GPG + \beta_2 Controls + \beta_3 Industry\ FE + \epsilon$, with controls for firm size (log market capitalization), the percent of female directors, and the firm's Asset4 environmental score. We include industry fixed effects based on 2-digit NAICS codes. $Good\ GPG = 1 - GPG$, where GPG is the mean hourly gender pay gap for the firm. The dependent variables are measures of voluntary disclosure related to gender diversity based on annual reports released in 2016. $GD\ disclosure$ is 1 if the firm discloses any of the voluntary disclosure categories. $\# Disclosure\ categories$ is the number of voluntary disclosure categories disclosed by the firm. $Word\ count$ is the number of words related to gender diversity. **Panel B** estimates the same regression as Panel A, except where $Disclosure$ is replaced by each of the disclosure categories. $Metrics$ is 1 if the firm discloses a quantitative gender diversity metric. $Targets$ is 1 if the firm discloses a gender diversity target with a quantitative goal and time-frame. $External\ commitments$ is 1 if the firm commits to an external gender diversity organization, such as the 30% Club. $Internal\ actions$ is 1 if the firm discloses an internal action targeted towards improving gender diversity, such as gender-blind recruiting. $Benefits$ is 1 if the firm discloses diversity-related benefits employees receive, such as offering flexible working. Standard errors are clustered at the industry level. t statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 4: Implication for ESG Ratings

<i>Panel A: Good GPG and Social Scores</i>			
	(1)	(2)	(3)
	Asset4	Sustainalytics	MSCI
Good GPG	-0.214** (-2.26)	0.055 (0.51)	0.011 (0.05)
Size	0.042*** (3.43)	-0.003 (-0.19)	-0.024 (-1.64)
Female directors	0.289 (1.75)	0.142 (0.75)	0.085 (0.36)
Environmental score	0.492*** (7.08)	0.304*** (4.28)	0.108 (1.68)
N	98	98	98
Adj. R-squared	0.531	0.162	0.172
Industry FE	Yes	Yes	Yes
Cluster	Industry	Industry	Industry
<i>Panel B: Voluntary Disclosure Categories and Social Scores</i>			
	(1)	(2)	(3)
	Asset4	Sustainalytics	MSCI
Metrics	0.004 (0.15)	-0.010 (-0.49)	0.010 (0.39)
Targets	0.112** (2.34)	0.064*** (3.05)	0.063 (1.36)
External commitments	0.116*** (3.15)	-0.015 (-0.52)	-0.029 (-0.51)
Internal actions	0.097** (2.66)	0.051* (1.95)	0.012 (0.57)
Benefits	0.007 (0.16)	0.050** (2.83)	0.076* (1.97)
N	167	167	167
Adj. R-squared	0.127	0.133	0.172
Adj. within R-squared	0.135	0.084	0.013
Industry FE	Yes	Yes	Yes
Cluster	Industry	Industry	Industry

This table presents the practical implication of our main result on ESG ratings. **Panel A** shows the results from estimating the regression $Social\ Score = \beta_0 + \beta_1 Good\ GPG + \beta_2 Controls + \beta_3 Industry\ FE + \epsilon$, with controls for firm size (log market capitalization), the percent of female directors, and the firm's environmental score from each rating provider. We include industry fixed effects based on 2-digit NAICS codes. $Good\ GPG = 1 - GPG$, where GPG is the mean hourly gender pay gap for the firm. The social scores are from Asset4, Sustainalytics, and MSCI. **Panel B** examines the relationship between the voluntary disclosure categories and the social scores. $Metrics$ is 1 if the firm discloses a quantitative gender diversity metric. $Targets$ is 1 if the firm discloses a gender diversity target with a quantitative goal and time-frame. $External\ commitments$ is 1 if the firm commits to an external gender diversity organization, such as the 30% Club. $Internal\ actions$ is 1 if the firm discloses an internal action targeted towards improving gender diversity, such as gender-blind recruiting. $Benefits$ is 1 if the firm discloses diversity-related benefits employees receive, such as offering flexible working. Standard errors are clustered at the industry level. t statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 5: Potential Explanations

<i>Panel A: Labor Market Information</i>	
	(1)
	The Times Top 50 Employer for Women
Medium GPG	0.090*** (3.69)
Better GPG	0.107 (1.74)
N	116
Adj. R-squared	0.236
Industry FE	Yes
Control	Yes
Cluster	Industry

Continued on following page

Table 5 (continued)

<i>Panel B: Selective Disclosure</i>		
	(1)	(2)
	Metrics: % Women in leadership	Metrics: % Women in leadership
Medium GPG	-0.207** (-2.98)	
Better GPG	-0.199** (-2.81)	
Worse GPG × % Female top pay quartile		0.825* (2.13)
Medium GPG × % Female top pay quartile		0.328 (0.93)
Better GPG × % Female top pay quartile		0.346 (1.71)
N	116	105
Adj. R-squared	0.090	0.048
F-test (W=M)		0.019
F-test (W=B)		0.123
F-test (M=B)		0.950
Industry FE	Yes	Yes
Control	Yes	Yes
Cluster	Industry	Industry

This table explores a potential explanation behind the negative relation between *Good GPG* and voluntary gender diversity disclosure. **Panel A** studies the relation of *Good GPG* and labor market information. Column 1 shows the result from estimating the regression $Y = \beta_0 + \beta_1 \text{Medium GPG} + \beta_2 \text{Better GPG} + \beta_3 \text{Controls} + \beta_4 \text{Industry FE} + \epsilon$ with controls for a firm's size (log market capitalization), percent of female directors, and Asset4 environmental score. *The Times Top 50 Employers for Women* is 1 if the firm received *The Times Top 50 Employers for Women* award in the UK before 2017. $\text{Good GPG} = 1 - \text{GPG}$, where *GPG* is the mean hourly gender pay gap for the firm; *Good GPG* is grouped in terciles within-industry. **Panel B** shows the result on selective disclosure. Column 1 estimates the same regression as Panel A except with the dependent variable of *Metric: % Women in leadership*, which is 1 if the firm discloses a voluntary gender diversity metric related to women in leadership. Column 2 replaces *Medium GPG* and *Better GPG* with each of the three within-industry terciles of *Good GPG* interacted with *% Female top pay quartile*, the percent of females in the firm's top pay quartile. Standard errors are clustered at the industry level. *t* statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Appendix

A1. Gender Composition Persistence

A2. Sample Representation

A3. GPG Determinants

A4: GPG (Absolute Value)

A5. 3-digit NAICS

A6: Median GPG

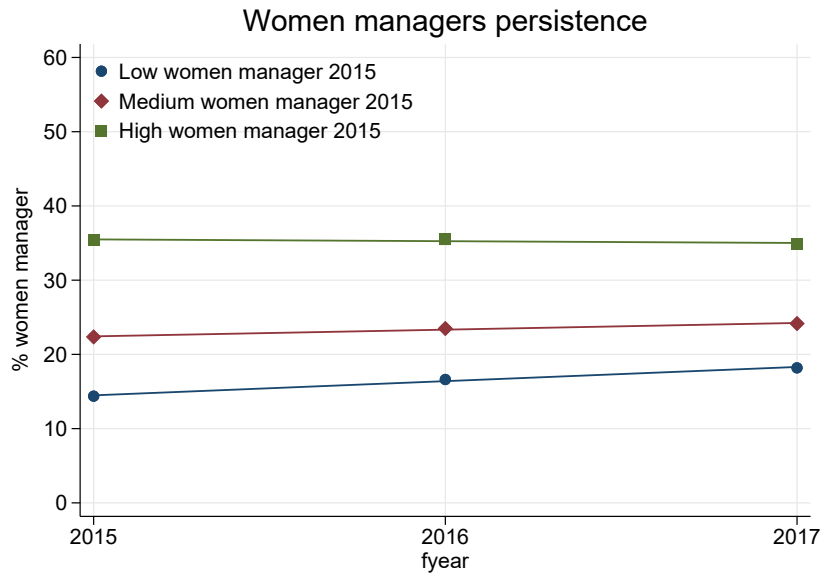
A7: Potential Explanation: Shareholder Pressure

A8: Potential Explanation: Room for Improvement

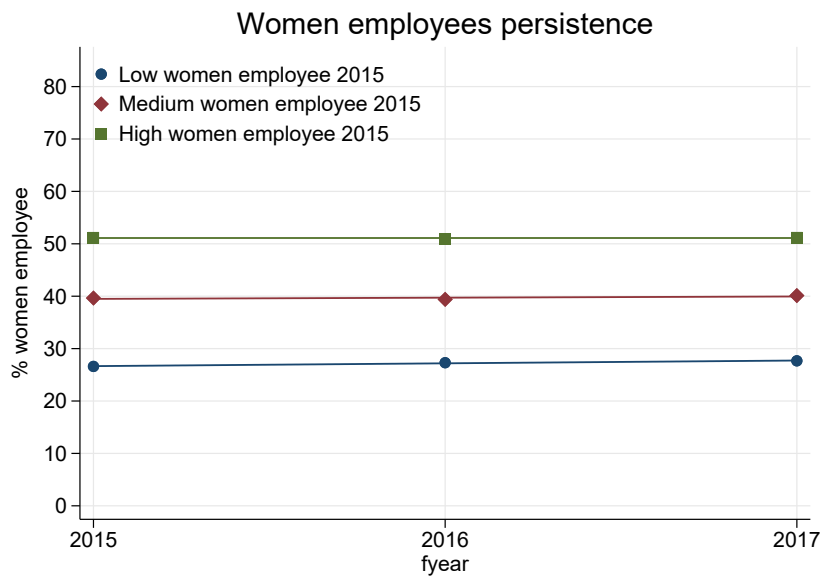
B. Data Collection Process

Figure A1: Gender Composition Persistence

Panel A: Female Senior Managers



Panel B: Female Employees



This figure shows the persistence of female composition in the firm between 2015 and 2017. Firms are grouped into tertiles within-industry based on their female composition in 2015.

Table A2: Sample Representativeness for Voluntary Disclosure Categories

	All			Sample		
	N	Mean	Std. dev.	N	Mean	Std. dev.
With GD disclosure	216	0.477	0.501	123	0.455	0.500
# disclosure categories	103	1.583	0.846	56	1.643	0.841
Word count	216	66.759	113.727	123	69.683	121.950
Targets	216	0.074	0.262	123	0.073	0.261
Metrics	216	0.287	0.453	123	0.244	0.431
External commitments	216	0.074	0.262	123	0.073	0.261
Internal actions	216	0.292	0.456	123	0.333	0.473
Benefits	216	0.125	0.331	123	0.130	0.338

This table shows the representativeness of our sample by comparing the rate of voluntary between all FTSE 350 firms with annual reports, and those in our sample used in the main analysis. For details of the reduction in sample size, see Table 1.

In Appendix A3, we shed some light on the determinants of *Good GPG*. In columns 1 and 2, we find that industry explains around a quarter of the variation in *Good GPG* and larger firms tend to have higher *Good GPG*. In column 3, we include the three mandatory gender composition metrics firms were required to report since 2013. Together, these three metrics explain 9.6% of the within-industry variation in *Good GPG*. This suggests our measure for *Good GPG* captures significant additional information compared to the three prior mandatory metrics. In column 4, we include the percentage of women in the four pay quartiles, which is also required in GPG reports. These more detailed composition metrics explain almost half of the within-industry variation in *Good GPG*. As expected, having a higher proportion of women in the two lowest pay quartiles is negatively associated with *Good GPG* and having a higher proportion of women in the top two pay quartiles is positively associated with *Good GPG*. In column 5, we include all determinant variables. Most observations remain similar, except the pay quartiles mostly subsume the explanatory power of the three mandatory gender composition metrics. After including the pay quartiles, though, 43% of the within-industry variation in *Good GPG* is still unexplained. These observations remain qualitatively similar when we use the smaller sample in column (5) for columns (1) through (4).

Table A3: Determinants of Good GPG

	<i>Good GPG</i>				
	(1)	(2)	(3)	(4)	(5)
Size		0.020*** (3.64)			0.021** (2.61)
% Female employees			-0.267** (-2.67)		-0.198 (-0.97)
% Female senior managers			0.048 (0.45)		-0.038 (-0.41)
% Female directors			0.162 (1.65)		-0.026 (-0.35)
% Female low pay quartile				-0.311* (-2.07)	-0.143 (-0.74)
% Female low-middle pay quartile				-0.402** (-2.72)	-0.525*** (-3.34)
% Female upper-middle pay quartile				0.103 (0.69)	0.243** (2.76)
% Female top pay quartile				0.466*** (3.98)	0.523*** (6.54)
N	123	117	98	112	87
Adj. R-squared	0.253	0.294	0.278	0.602	0.638
Adj. within R-squared	0.000	0.049	0.096	0.477	0.573
Industry FE	Yes	Yes	Yes	Yes	Yes
Cluster	Industry	Industry	Industry	Industry	Industry

t statistics in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

This table presents the determinants of *Good GPG*. $Good\ GPG = 1 - GPG$, where *GPG* is the mean hourly gender pay gap for the firm. All columns include industry fixed effects. Column 2 includes size, which is the log of market capitalization. Column 3 includes three female composition metrics that were required for UK firms since 2013. Column 4 includes female composition in the four pay quartiles, which was first reported as part of a firm's GPG disclosure for April 2017. Column 5 includes all explanatory variables from the preceding columns.

Table A4: Voluntary Disclosure and Good GPG (absolute value)

	(1)	(2)	(3)
	GD disclosure	# Disclosure categories	Word count
Good GPG	-1.700*** (-3.65)	-2.414* (-1.91)	-220.701 (-1.66)
Size	0.026 (0.63)	0.142 (1.46)	20.441 (1.31)
Female directors	1.213** (2.70)	2.886*** (3.78)	67.802 (1.23)
Environmental score	0.034 (0.19)	0.372 (1.32)	39.872 (0.56)
N	116	116	116
Adj. R-squared	0.141	0.218	0.137
Industry FE	Yes	Yes	Yes
Control	Yes	Yes	Yes
Cluster	Industry	Industry	Industry

t statistics in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

This table repeats the main analysis using the absolute value of GPG instead of the signed value of GPG to calculate *Good GPG*. It shows the results from estimating the regression $Disclosure = \beta_0 + \beta_1 Good\ GPG + \beta_2 Controls + \beta_3 Industry\ FE + \epsilon$ with controls for a firm's size (log market capitalization), percent of female directors, and Asset4 environmental score. Here only, *Good GPG* is $1 - |GPG|$, where *GPG* is the mean hourly gender pay gap for the firm. The dependent variables are measures of voluntary disclosure related to gender diversity based on annual reports released in 2016. *GD disclosure* is 1 if the firm discloses any of the voluntary disclosure categories. *# Disclosure categories* is the number of voluntary disclosure categories disclosed by the firm. *Word count* is the number of words related to gender diversity.

Table A5: Voluntary Disclosure and Good GPG within 3-digit-NAICS

	(1)	(2)	(3)
	GD disclosure	# Disclosure categories	Word count
Good GPG	-1.899*** (-4.14)	-3.239** (-2.70)	-433.669*** (-3.08)
Size	0.026 (0.61)	0.153 (1.50)	35.765 (1.51)
Female directors	1.239** (2.20)	3.469*** (3.09)	115.330 (1.12)
Environmental score	0.022 (0.07)	0.211 (0.31)	-60.279 (-0.50)
N	95	95	95
Adj. R-squared	0.166	0.265	0.121
Industry FE	Yes	Yes	Yes
Control	Yes	Yes	Yes
Cluster	Industry	Industry	Industry

t statistics in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

This table repeats the main analysis with 3-digit-NAICS as the industry classification instead of the 2-digit-NAICS used in the main analysis. It shows the results from estimating the regression $Disclosure = \beta_0 + \beta_1 Good\ GPG + \beta_2 Controls + \beta_3 Industry\ FE + \epsilon$ with controls for a firm's size (log market capitalization), percent of female directors, and Asset4 environmental score. $Good\ GPG = 1 - GPG$, where GPG is the mean hourly gender pay gap for the firm. The dependent variables are measures of voluntary disclosure related to gender diversity based on annual reports released in 2016. $GD\ disclosure$ is 1 if the firm discloses any of the voluntary disclosure categories. $\# Disclosure\ categories$ is the number of voluntary disclosure categories disclosed by the firm. $Word\ count$ is the number of words related to gender diversity.

Table A6: Voluntary Disclosure and Median GPG

	(1)	(2)	(3)
	GD disclosure	# Disclosure categories	Word count
Good GPG	-0.712 (-1.34)	-1.097 (-1.11)	-126.920 (-1.54)
Size	0.002 (0.03)	0.113 (0.94)	11.224 (1.00)
Female directors	1.267** (2.48)	2.977*** (3.88)	97.138 (1.57)
Environmental Score	0.041 (0.20)	0.320 (0.88)	58.648 (0.91)
N	114	114	114
Adj. R-squared	0.047	0.173	0.157
Industry FE	Yes	Yes	Yes
Control	Yes	Yes	Yes
Cluster	Industry	Industry	Industry

t statistics in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

This table repeats the main analysis using the median GPG instead of the mean GPG. It shows the results from estimating the regression $Disclosure = \beta_0 + \beta_1 Good\ GPG + \beta_2 Controls + \beta_3 Industry\ FE + \epsilon$ with controls for a firm's size (log market capitalization), percent of female directors, and Asset4 environmental score. $Good\ GPG = 1 - GPG$, where GPG is the median hourly gender pay gap for the firm. The dependent variables are measures of voluntary disclosure related to gender diversity based on annual reports released in 2016. $GD\ disclosure$ is 1 if the firm discloses any of the voluntary disclosure categories. $\# Disclosure\ categories$ is the number of voluntary disclosure categories disclosed by the firm. $Word\ count$ is the number of words related to gender diversity.

One potential explanation of our main result centers on investor pressure of firms with poor gender diversity. In the period before GPG Mandate, investors may have received information allowing them to infer that certain firms had poor gender diversity. Then investors may have pressured those firms to improve, leading the firms to voluntarily disclose more about their gender diversity efforts.²¹ As a result, this investor pressure could explain why we see firms with worse GPGs be more likely to provide voluntary disclosure.

To examine this potential explanation, first we read news articles about investors pressuring UK firms over gender diversity in 2014 and 2015. We find anecdotal evidence that investors pressure firms with poor gender composition on their boards. For example, investors such as Old Mutual Global Investors and Hermes Equity Ownership Services announced they would likely reject board proposals at annual meetings of FTSE 250 firms with all-male boards ([Marriage, 2015](#)). The business press also named and shamed FTSE 350 firms with relatively low women on their boards such as JD Sports Fashion and Persimmon. Based on this review, poor board gender composition was likely the primary driver behind investor pressure at the time. Because quoted firms had to disclose the gender composition of their boards, senior managers, and total employees since 2013, it is reasonable investors would use that publicly available information to infer which firms to pressure.

Because investors were pressuring firms with worse board gender composition, we examine whether these firms were more likely to disclose voluntarily. In Appendix A7, we repeat our main analysis but replace *Good GPG* with an indicator for firms with less than 25% women on their boards, the recommended ratio based on the Davies report ([Davies, 2021](#)). We do not find evidence that firms with worse board gender composition are more likely to voluntarily disclose; if anything, firms with worse board gender composition appear less likely to voluntarily disclose but the coefficients are not statistically significant.

²¹This reasoning is also similar to the legitimacy theory, where firms subject to investor pressure use disclosure as a legitimizing device to change the public perception of the firm ([Patten, 2002](#)).

Table A7: Potential Explanation: Investor pressure

	(1) GD disclosure	(2) # Disclosure categories	(3) Word count
< 25% Female directors	-0.158 (-1.16)	-0.288 (-1.09)	-5.158 (-0.24)
Size	-0.001 (-0.02)	0.121 (0.73)	16.449 (1.17)
Environmental score	0.145 (0.73)	0.612 (1.62)	52.718 (0.75)
N	116	116	116
Adj. R-squared	0.009	0.119	0.112
Industry FE	Yes	Yes	Yes
Control	Yes	Yes	Yes
Cluster	Industry	Industry	Industry

t statistics in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

This table tests whether firms with fewer female board directors were more likely to voluntarily disclose about gender diversity. The table repeats the main analysis in Table 2 but replaces *Good GPG* with an indicator for firms with less than 25% female directors in 2015. We control for firm size (log market capitalization) and Asset4 environmental score. The dependent variables are measures of voluntary disclosure related to gender diversity based on annual reports released in 2016. *GD disclosure* is 1 if the firm discloses any of the voluntary disclosure categories. *# Disclosure categories* is the number of voluntary disclosure categories disclosed by the firm. *Word count* is the number of words related to gender diversity.

Another potential explanation is that firms with better GPGs are less likely to disclose because they have less room for improvement. We investigate this explanation using the disclosure of targets: intuitively a target represents an improvement goal, and a firm may not need to set a target if it is already performing relatively well on gender diversity. As demonstrated in Table 2 Panel B, firms with *Medium GPG* are more likely to disclose targets. Among the eight firms with targets, two are *Better GPG* firms, five are *Medium GPG* firms, and one is a *Worse GPG* firm. Because targets are relatively rare, we examine each firm closely to determine whether room for improvement motivates the target disclosure.

Appendix Table A8 summarizes each firm's target disclosure. Out of the eight firms, five set targets consistent with the room-for-improvement explanation. These targets include reaching a specific ratio of female representation related to senior leadership or new recruits. One example is SSE PLC, whose target is for 25% of their female employees to earn at least £40,000 by 2025. While this is a rather unusual target, it captures the dimension where SSE has more room for improvement: in 2015, only 10% of SSE's female employees earned at least £40,000. By contrast, SSE did not set a senior management target and 30% of their senior managers were women, above average in their industry.

Interestingly, the two *Better GPG* firms that disclose targets have targets that are unrelated to their employees. Diageo PLC and Unilever PLC set a target to "empower" 2 million and 5 million women by 2016 and 2020, respectively. We hesitate to extrapolate from two cases, but the fact that each firm has less room to improve on traditional manager metrics and that each firm's target is relatively unrelated to their workforce suggests that in general, firms with better GPGs do not disclose targets because they have less room to improve. Inversely, the anecdotes suggest firms with worse GPGs may be more likely to disclose because they have more room for improvement.

Table A8: Gender Diversity Targets

Firm	GPG	Target	Room to Improve	Comment
Diageo PLC	Better	Empower 2 million women	No	High performer on multiple gender diversity dimensions, hence target not related to workforce.
Unilever PLC	Better	Empower 5 million women	No	High performer on multiple gender diversity dimensions, hence target not related to workforce.
Astrazeneca	Medium	41% Senior Management	No	High performer on women management dimension, target is less about room for improvement.
Kier Group PLC	Medium	30% New Recruits	Yes	While the target is specific and relates to room for improvement, it is less costly to achieve than a senior manager target.
Lloyds Banking Group PLC	Medium	40% Senior Management	Yes	While percent of women managers higher than industry average, there is still room for improvement. Also in an industry in the media spotlight for gender diversity concerns.
WM Morrison Supermarkets PLC	Medium	30% Senior Management	Yes	Has a lower percent of women managers than industry average, so there is room for improvement. Yet this is a missed target.
Pennon Group PLC	Medium	25% Senior Management	Yes	Has a lower percent of women managers than industry average, so there is room for improvement.
SSE PLC	Worse	25% over £40,000	Yes	Has a higher percent of women managers than industry average. While the target is specific and unusual, it reflects room for improvement since the percent of women managers already pretty high.

This table lists and describes the firms with gender diversity targets, ranked by GPG terciles. This table provides support for the room-for-improvement channel.

Appendix B: Data Collection Process

Pilot Analysis

We randomly select 10 firms among the FTSE 100 as of June 2019 to run a pilot data collection. For fiscal years before 2017, we manually read the annual reports and CSR or sustainability reports for any discussion related to gender diversity, and save the relevant text into a word document.

The pilot analysis provides the following four insights that shape our final data collection process. First, we observe five common types of disclosure: (i) metrics, (ii) targets, (iii) external commitments, (iv) internal actions, and (v) policy benefits. Second, we observe many mentions of the GPG Mandate regulation in annual reports released in 2017, so we decide to use reports released in 2016 to mitigate anticipation concerns. Third, we observe that when firms disclose about gender diversity, they tend to do so in their annual reports, and their sustainability reports tend to mimic the gender diversity text in the annual reports. As a result, we decide to focus on voluntary disclosure in annual reports, as opposed to sustainability reports. Finally, we observe that firms tend to discuss gender diversity in two separate sections: one section about the firm's employees or culture and another section about the board nomination process. Because we are interested in a firm's voluntary gender diversity disclosure and board diversity discussion is closely tied to a pre-existing mandatory reporting requirement, we exclude any disclosure related to the board when we collect voluntary disclosure.

File collection

Based on insights from the pilot analysis, we expand the sample to the FTSE 350 firms and collect their fiscal 2015 annual reports released in 2016 and 2017 GPG reports. We hand collect the data for FTSE 350 with help from a research assistant. Annual reports are downloaded from corporate websites or <https://www.annualreports.com/>. GPG reports are downloaded from corporate websites; where we cannot find the GPG report on the firm's website, we use the links provided in the UK government's gender pay gap data (see section 3.2).

Data collection

Company GPG data

We collect two pieces of information from a firm's GPG report. First, we record the number of subsidiaries for which the firm is required to disclose GPG data; we cross-reference these subsidiaries with the UK government GPG dataset. During this process, we note whether the parent entity or consolidated firm is also required to disclose under GPG Mandate. Second, when the consolidated firm reports GPG metrics for its UK workforce, we record these metrics.

Voluntary disclosure data

For each firm, we repeat the following process.

Step 1: We identify sections in the annual report that relate to gender diversity but are unrelated to corporate boards. We search for the following keywords to identify the relevant sections: female, woman, women, gender, diversity, inclusion. Additionally, we search for specific keywords that appeared in the pilot analysis in the context of gender diversity: 30% club; unconscious bias; pipeline; mentorship; matern; patern; parental; flexible work. We save the relevant text into a word document for further processing, and record the number of words in the document as a measure of disclosure.

Step 2: Using the word document, we group the text into the five common categories.

Metrics includes gender diversity metrics such as the percentage of women in each division or the percentage of women earning above a certain salary. Metrics does not include the percentage of women among the board, senior managers, or total employees because those specific data points were already required to be disclosed. Targets include quantitative gender diversity targets with a target year. External commitments includes commitments to external gender diversity organizations. Internal actions includes internal actions the firm is taking to improve gender diversity. Finally, policy benefits includes policies that firms explicitly link to encouraging gender diversity. We create a binary variable for each disclosure category so that if a firm discloses anything in a category, we code that category as 1.

We perform two procedures to validate the consistency of the data collection. First, we each cross-check 10 samples collected by another person to make sure we collect and categorize the data consistently. Second, we keep a list of text where the categorization is uncertain, and we discuss the classification together to ensure consistent application of category definitions. Below are three examples of uncertain cases.

Example 1: “Our Global Diversity Policy makes clear the responsibility of all employees and workers to treat colleagues with dignity and respect and to create an inclusive environment free from discrimination, bullying, harassment or victimisation, irrespective of their age, colour, disability, ethnic or national origin, gender, gender identity/expression, marital status, pregnancy, race, religion or belief, or sexual orientation.”

This paragraph is excluded from our data collection because it is too general and only weakly relate to gender.

Example 2: “Our seven global employee networks support this strategy and focus on gender, age, ethnicity, LGBT+, faith, working parents and carers, and ability. We have continued our focus on improving gender balance within senior leadership.”

This paragraph is included in our data collection because the first sentence tells us there is a gender-based employee network, which is a specific action, and the second sentence explicitly touches on gender balance among senior leadership.

Example 3: “Among the top 101 executive managers, 23 (23%) were women compared with 18% in 2014.”

This paragraph is included in our data collection and is classified as metrics. While firms are required to disclose the percentage of women among senior managers, measures like “top 101 executive managers” are not covered by the standard definition of senior managers. Thus discussing the “top 101 executive managers” is not mandatory, and we treat this sentence as a voluntary disclosure of metrics.