Online Appendix

Pay Transparency and Gender Equality

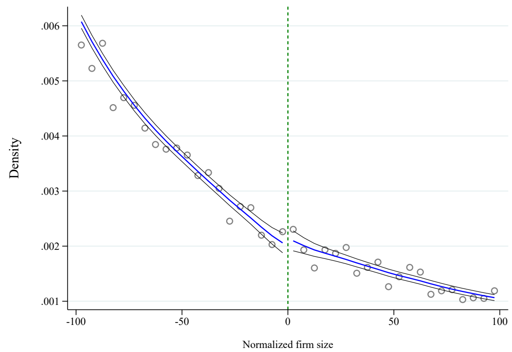
Jack Blundell Emma Duchini Ştefania Simion Arthur Turrell

A Further information, results and robustness checks

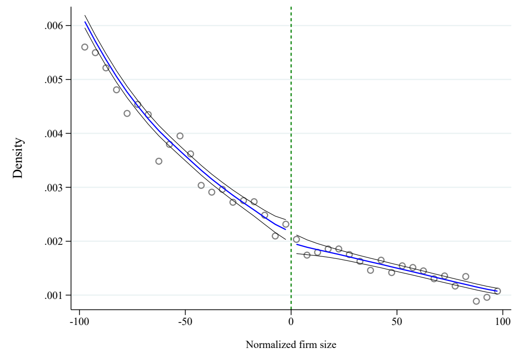
A.1 Business Structure Database

When none of the employees of a firm is interviewed in ASHE in the year used to define the treatment status, we recover the information on firm size from the Business Structure Database (BSD). BSD provides information on firm output, employment, and turnover for almost 99 percent of business organizations registered in the UK. The data come from the Inter-Departmental Business Register (IDBR), a live register of firms collected by the tax authorities via VAT and employee tax records. ASHE and BSD provide the same anonymized firm identifier which allows us to match them with each other. Importantly, when merging the two data sets by firm and year, we merge ASHE data for a specific year with BSD data for the previous year. This is because in 74 percent of cases in which both data sets have non-missing information on number of employees, ASHE number of employees in a specific year coincides with BSD number of employees for the previous year, while in only 40 percent of cases, ASHE number of employees coincides with BSD number of employees for the same year. Moreover, when ASHE number of employees differs from BSD number of employees for the previous year, the average difference is only 1.7 employees. In contrast, when ASHE number of employees differs from BSD number of employees for the same year, the average difference is 6.3 employees. From conversations with the ONS, it appears that this time discrepancy between the two data sets is probably due to the time of the year in which the information on firms' number of employees is collected. Importantly, Table A9 shows that our results are practically unchanged when using BSD number of employees for the same year to perform this imputation.

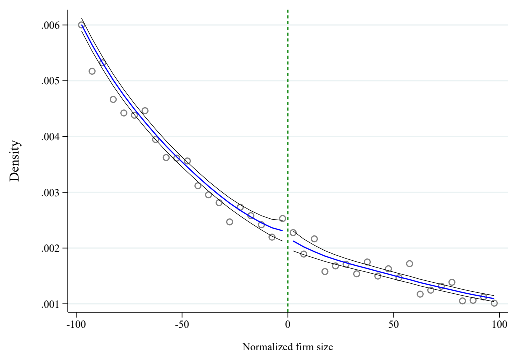
Figure A1: Firm size distribution



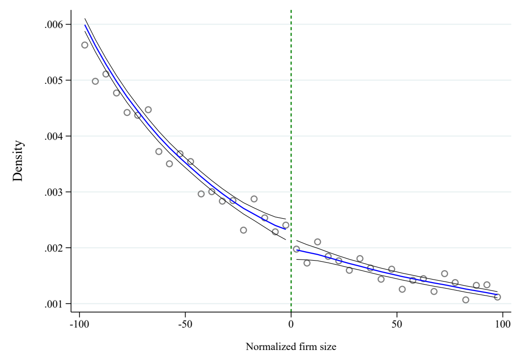
(A) BSD 2016



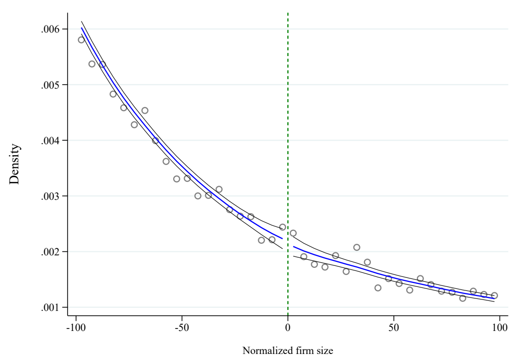
(B) BSD 2017



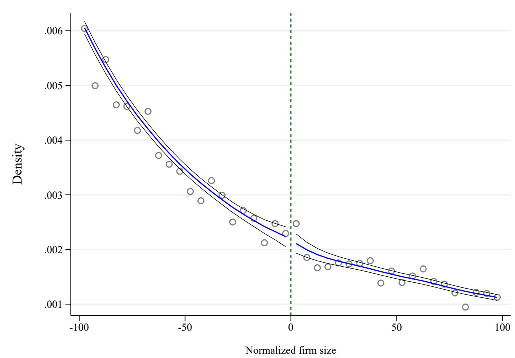
(C) BSD 2018



(D) BSD 2019



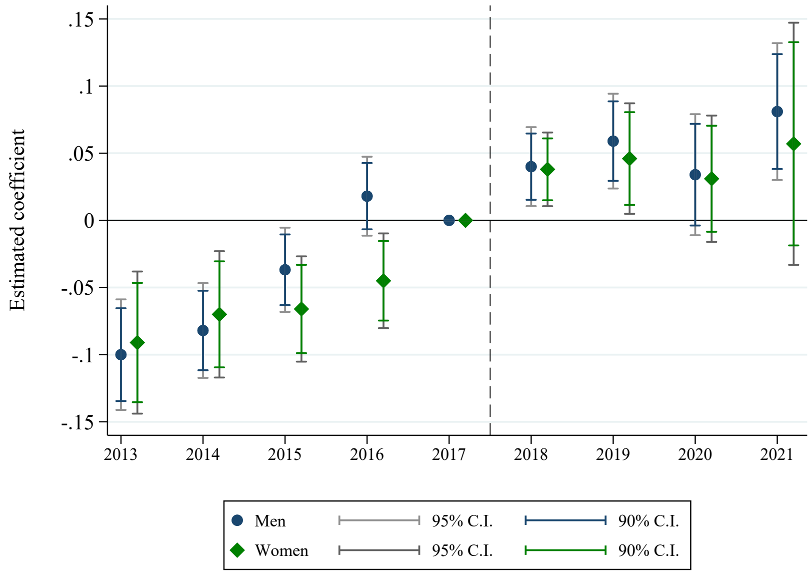
(E) BSD 2020



(F) BSD 2021

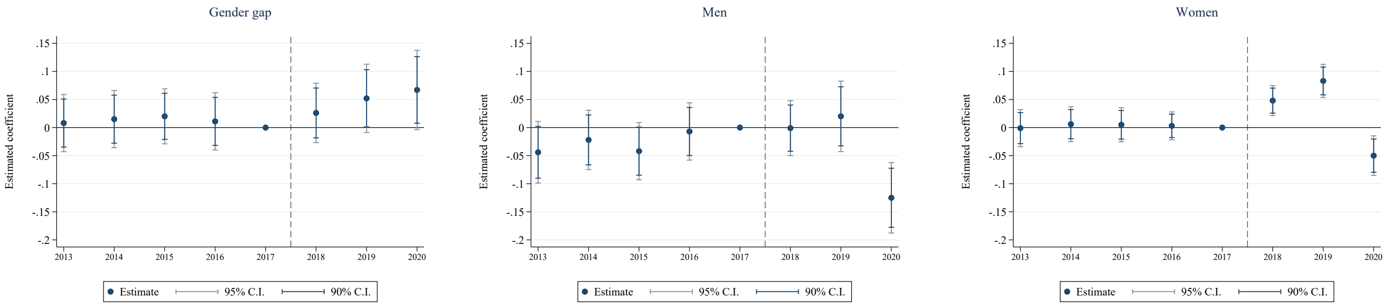
Source: BSD, 2016–2021.
Note: These graphs show the distribution of firms around the 250-employee cutoff in each year since the announcement of the policy. In each figure, the sample includes firms with +/100 employees from the threshold, grouped in 20 bins. Each dot represents the share of firms with a number of employees comprised in the corresponding bin.

Figure A2: General equilibrium effects



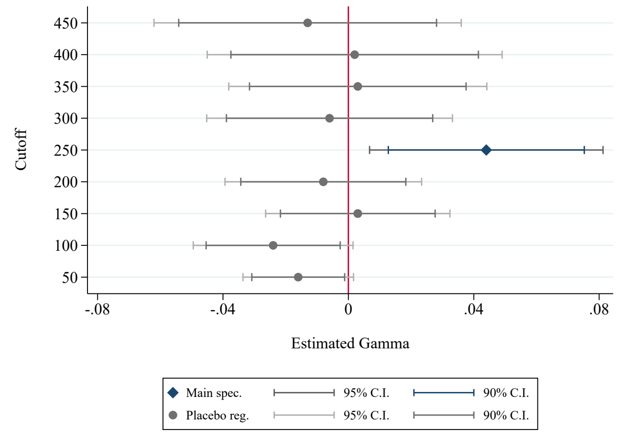
Source: ASHE, 2013–2021.
Notes: These graphs present the estimates of the year-specific effects for male and female workers employed in control firms. These results are obtained from the estimation of regression 1. The estimation sample includes workers employed in firms with 200 to 300 employees. The graphs also report 90 and 95 percent confidence intervals associated with firm-level clustered standard errors. The dash vertical line indicates the month when the mandate is approved, i.e., February 2017.

Figure A3: Event studies - separations

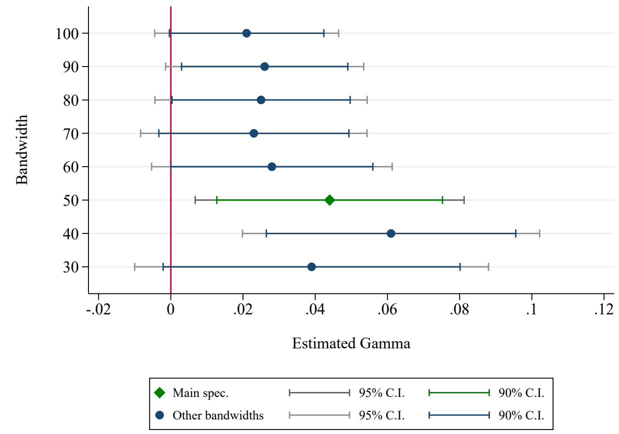


Source: ASHE, 2013–2020.
Notes: These graphs present the estimates of the leads and lags of the policy on workers' separations. These results are obtained from the estimation of regression 2, using firm fixed effects in place of firm times individual fixed effects. In each graph, the estimation sample includes workers employed in firms with 200 to 300 employees. The graphs also report 90 and 95 percent confidence intervals associated with firm-level clustered standard errors. The dash vertical line indicates the month when the mandate is approved, i.e., February 2017.

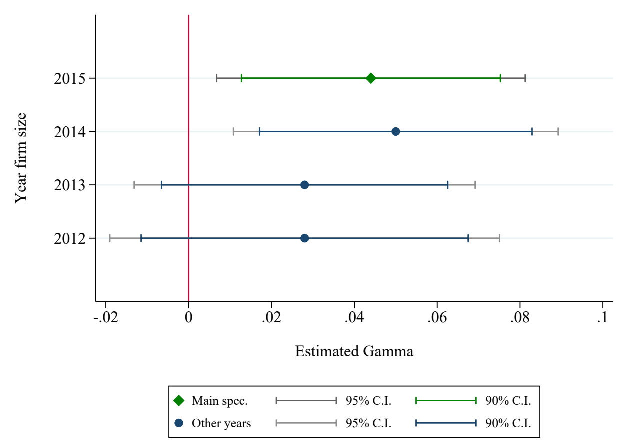
Figure A4: Robustness checks - gender gap in separations



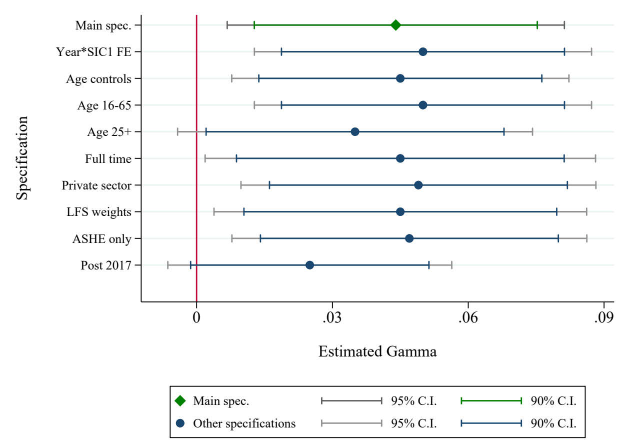
(A) Placebo regressions



(B) Changing bandwidth



(C) Changing year treatment status



(D) Other robustness checks

Source: ASHE, 2013–2020.
Notes: These graphs present a series of robustness checks on the impact of the policy on the gender gap in separations.

Table A1: Event studies - log hourly pay

	Gender pay gap (1)	Men (2)	Women (3)
Effect 2013	0.011	0.011	0.005
	(0.012)	(0.012)	(0.017)
Effect 2014	0.015	0.015	0.003
	(0.011)	(0.011)	(0.016)
Effect 2015	0.012	0.012	-0.013
	(0.009)	(0.009)	(0.015)
Effect 2016	0.014	0.014	0.005
	(0.009)	(0.009)	(0.013)
Effect 2018	-0.007	-0.007	0.011
	(0.008)	(0.008)	(0.014)
Effect 2019	-0.036***	-0.036***	-0.007
	(0.010)	(0.010)	(0.015)
Effect 2020	-0.023	-0.023	-0.011
	(0.017)	(0.017)	(0.018)
Effect 2021	-0.034*	-0.034*	-0.005
	(0.019)	(0.019)	(0.024)
Add Effect Fem 2013	-0.006		
	(0.020)		
Add Effect Fem 2014	-0.012		
	(0.019)		
Add Effect Fem 2015	-0.025		
	(0.017)		
Add Effect Fem 2016	-0.009		
	(0.015)		
Add Effect Fem 2018	0.018		
	(0.015)		
Add Effect Fem 2019	0.030*		
	(0.017)		
Add Effect Fem 2020	0.012		
	(0.023)		
Add Effect Fem 2021	0.029		
	(0.030)		
Observations	35,092	18,871	16,221
Adjusted R^2	0.894	0.919	0.856

Source: ASHE, 2013–2021.

Notes: This table presents event-study estimates of the effect of the pay transparency policy on the log hourly pay. The results in Column 1 are obtained from the estimation of regression 2, while results in Column 2 and 3 are obtained from the estimation of the difference-in-difference analogue of this specification for each gender. The estimation sample comprises men and women working in firms that have between 200 and 300 employees. All regressions include firm*individual fixed effects and gender-region specific time shocks. A treated firm is defined as having at least 250 employees in 2015. The post dummy is equal to one from 2018 onward. Heteroskedasticity-robust standard errors clustered at firm level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table A2: Impact on pay and hours worked

	Log hourly pay (1)	Log weekly pay (2)	Weekly hours (3)	Part-time (4)
Treated firm*post	-0.029***	-0.016	0.223	-0.008
	(0.009)	(0.011)	(0.191)	(0.009)
Treated firm*post*fem	0.030**	0.008	-0.515	0.028*
	(0.013)	(0.020)	(0.377)	(0.017)
Observations	35,092	35,092	35,092	35,092
Adjusted R^2	0.894	0.904	0.789	0.744
P-value Women Coeff	0.909	0.632	0.370	0.250
Men's pre-policy mean	15.94	581.73	36.41	0.10
Women's pre-policy mean	13.36	414.52	30.69	0.34

Source: ASHE, 2013–2021.

Notes: This table reports the impact of pay transparency on pay outcomes and hours worked, obtained from the estimation of regression 1. Each column refers to a different outcome, as specified at the top of it. The estimation sample comprises men and women working in firms that have between 200 and 300 employees. All regressions include firm*individual fixed effects and gender-region specific time shocks. A treated firm is defined as having at least 250 employees in 2015. The post dummy is equal to one from 2018 onward. Heteroskedasticity-robust standard errors clustered at firm level in parentheses. The p-value at the bottom of the table refers to the t-test on the sum of the two reported coefficients, corresponding to the effect of the policy on female employees. The pre-policy mean represents the mean of the outcome variable for the treated group between 2013 and 2017.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A3: Impact on pay cuts

	Nominal pay cut (1)	Real pay cut (2)
Treated firm*post	-0.015	-0.012
	(0.016)	(0.022)
Treated firm*post*fem	-0.003	-0.008
	(0.023)	(0.030)
Observations	35,092	35,092
Adjusted R^2	0.003	0.072
P-value Women Coeff	0.341	0.381
Men's pre-policy mean	0.12	0.32
Women's pre-policy mean	0.13	0.30

Source: ASHE, 2013–2021.

Notes: This table reports the impact of pay transparency on pay cuts, obtained from the estimation of regression 1. Each column refers to a different outcome, as specified at the top of it. The estimation sample comprises men and women working in firms that have between 200 and 300 employees. All regressions include firm*individual fixed effects and gender-region specific time shocks. A treated firm is defined as having at least 250 employees in 2015. The post dummy is equal to one from 2018 onward. Heteroskedasticity-robust standard errors clustered at firm level in parentheses. The p-value at the bottom of the table refers to the t-test on the sum of the two reported coefficients, corresponding to the effect of the policy on female employees. The pre-policy mean represents the mean of the outcome variable for the treated group between 2013 and 2017.

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

Table A4: Impact on pay outcomes by occupation

	Entire sample (1)	Lower-paid occupations (2)	Better-paid occupations (3)	P-value T-test (4)
Treated firm*post	-0.029***	-0.021**	-0.040***	0.293
	(0.009)	(0.010)	(0.015)	
Treated firm*post*fem	0.030**	0.020	0.055**	0.463
	(0.013)	(0.017)	(0.022)	
Observations	35,092	20,002	14,476	
Adjusted R^2	0.894	0.745	0.890	
P-value Women Coeff	0.909	0.926	0.387	
Men's pre-policy mean	15.94	10.51	23.53	
Women's pre-policy mean	13.36	9.73	18.87	

Source: ASHE, 2013–2021.

Notes: This table compares the impact of pay transparency on employees' hourly pay across occupations, by estimating regression 1 by subgroup. Column 1 reports the estimate on log hourly pay for the entire sample, employees working in firms that have between 200 and 300 employees. Columns 2 and 3 compare the impact across the lower-paid and higher-paid occupations, where this grouping is based on the ranking of pre-policy 1-digit SOC-specific median wages. Column 4 reports the p-value of the t-test on the equality of estimates in Columns 2 and 3. All regressions include firm*individual fixed effects and gender-region specific time shocks. A treated firm is defined as having at least 250 employees in 2015. The post dummy is equal to one from 2018 onward. Heteroskedasticity-robust standard errors clustered at firm level in parentheses. The p-value at the bottom of the table refers to the t-test on the sum of the two reported coefficients, corresponding to the effect of the policy on female employees. The pre-policy mean represents the mean of the outcome variable for the treated group and subgroup considered between 2013 and 2017.

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

Table A5: Impact on log hourly pay - placebo regressions

	50	100	150	200	**250**	300	350	400	450
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treated firm*post	-0.004	-0.011**	-0.002	-0.004	-0.029***	-0.004	-0.011	-0.000	-0.016
	(0.004)	(0.006)	(0.007)	(0.008)	(0.009)	(0.010)	(0.011)	(0.012)	(0.015)
Treated firm*post*fem	-0.004	0.007	-0.002	0.001	0.030**	-0.008	0.019	0.013	0.024
	(0.006)	(0.009)	(0.010)	(0.013)	(0.013)	(0.015)	(0.018)	(0.018)	(0.021)
Observations	288,721	101,571	62,208	46,749	35,092	27,914	23,716	19,648	16,757
Adjusted R^2	0.845	0.883	0.890	0.892	0.894	0.899	0.898	0.892	0.895
P-value Women Coeff	0.109	0.523	0.682	0.709	0.909	0.344	0.591	0.364	0.606
Men's pre-policy mean	14.87	15.36	15.81	15.71	15.94	15.66	16.04	16.04	16.05
Women's pre-policy mean	11.99	12.79	12.87	13.49	13.36	13.53	13.04	13.50	13.06

Source: ASHE, 2013–2021.

Notes: This table reports the impact of placebo policies on log hourly pay, obtained from the estimation of regression 1. In each regression, the estimation sample comprises employees working in firms that have +/- 50 employees from the threshold c specified at the top of each column. All regressions include firm*individual fixed effects and gender-region specific time shocks. A treated firm is defined as having at least 250 employees in 2015. The post dummy is equal to one from 2018 onward. Heteroskedasticity-robust standard errors clustered at firm level in parentheses. The p-value at the bottom of the table refers to the t-test on the sum of the two reported coefficients, corresponding to the effect of the policy on female employees. The pre-policy mean represents the mean of the outcome variable for the treated group between 2013 and 2017.

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

Table A6: Impact on log hourly pay - different bandwidths

	30 (1)	40 (2)	**50** (3)	60 (4)	70 (5)	80 (6)	90 (7)	100 (8)
Treated firm*post	-0.027**	-0.024**	-0.029***	-0.024***	-0.026***	-0.027***	-0.021***	-0.018***
	(0.012)	(0.010)	(0.009)	(0.008)	(0.007)	(0.007)	(0.007)	(0.006)
Treated firm*post*fem	0.046**	0.034**	0.030**	0.020*	0.019*	0.020*	0.015	0.007
	(0.018)	(0.015)	(0.013)	(0.012)	(0.011)	(0.011)	(0.010)	(0.009)
Observations	19,291	27,257	35,092	43,154	51,713	60,208	69,126	78,702
Adjusted R^2	0.894	0.896	0.894	0.892	0.894	0.894	0.893	0.893
P-value Women Coeff	0.189	0.426	0.909	0.698	0.424	0.396	0.478	0.125
Men's pre-policy mean	16.09	16.14	15.94	15.78	15.87	15.88	15.84	15.83
Women's pre-policy mean	13.37	13.37	13.36	13.37	13.45	13.45	13.43	13.40

Source: ASHE, 2013–2021.

Notes: This table reports the impact of pay transparency on log hourly pay, obtained from the estimation of regression 1. In each regression, the estimation sample comprises individuals working in firms that have +/- h employees from the 250-employee threshold, where h is indicated at the top of each column. All regressions include firm*individual fixed effects and gender-region specific time shocks. A treated firm is defined as having at least 250 employees in 2015. The post dummy is equal to one from 2018 onward. Heteroskedasticity-robust standard errors clustered at firm level in parentheses. The p-value at the bottom of the table refers to the t-test on the sum of the two reported coefficients, corresponding to the effect of the policy on female employees. The pre-policy mean represents the mean of the outcome variable for the treated group between 2013 and 2017.

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

Table A7: Impact on log hourly pay - changing year to define treatment status

	Main spec (1)	Firm size 2014 (2)	Firm size 2013 (3)	Firm size 2012 (4)
Treated firm*post	-0.029***	-0.020**	-0.023**	-0.033***
	(0.009)	(0.009)	(0.009)	(0.012)
Treated firm*post*fem	0.030**	0.018	0.025*	0.037**
	(0.013)	(0.014)	(0.014)	(0.017)
Observations	35,092	34,787	34,444	25,934
Adjusted R^2	0.894	0.893	0.894	0.895
P-value Women Coeff	0.909	0.869	0.899	0.745
Men's pre-policy mean	15.94	15.80	16.09	16.22
Women's pre-policy mean	13.36	13.43	13.40	13.24

Source: ASHE, 2013–2021.

Notes: This table reports the impact of pay transparency on log hourly pay, obtained from the estimation of regression 1. In each regression, the estimation sample comprises men and women working in firms that have between 200 and 300 employees. All regressions include firm*individual fixed effects and gender-region specific time shocks. A treated firm is defined as having at least 250 employees in 2015 or in the year indicated on top of each column. The post dummy is equal to one from 2018 onward. Heteroskedasticity-robust standard errors clustered at firm level in parentheses. The p-value at the bottom of the table refers to the t-test on the sum of the two reported coefficients, corresponding to the effect of the policy on female employees. The pre-policy mean represents the mean of the outcome variable for the treated group between 2013 and 2017.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A8: Impact on log hourly pay - other robustness checks

	Main spec (1)	1-digit SIC FE (2)	Age controls (3)	25 + (4)	16-65 (5)	Private sector (6)	Full-time (7)	LFS weights (8)	ASHE only (9)	Post 2017 (10)
Treated firm*post	-0.029***	-0.029***	-0.026***	-0.029***	-0.022**	-0.023***	-0.030***	-0.031***	-0.027***	-0.026***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.009)	(0.009)
Treated firm*post*fem	0.030**	0.028**	0.029**	0.028**	0.027**	0.025*	0.027*	0.032**	0.031**	0.027**
	(0.013)	(0.014)	(0.013)	(0.014)	(0.013)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Observations	35,092	35,082	35,092	34,304	32,044	27,664	30,004	35,092	30,346	35,092
Adjusted R^2	0.894	0.894	0.896	0.894	0.901	0.925	0.894	0.900	0.899	0.894
P-value Women Coeff	0.909	0.931	0.793	0.934	0.633	0.887	0.766	0.889	0.748	0.959
Men's pre-policy mean	15.94	15.94	15.94	15.98	16.84	16.49	15.88	17.07	16.03	15.80
Women's pre-policy mean	13.36	13.36	13.36	13.39	14.05	14.05	13.03	13.88	13.36	13.35

Source: ASHE, 2013–2021.

Notes: This table reports a series of robustness checks on the impact of pay transparency on log hourly pay, obtained from the estimation of regression 1. In each regression, the estimation sample comprises men and women working in firms that have between 200 and 300 employees. A treated firm is defined as having at least 250 employees in 2015. The post dummy is equal to one from 2018 onward. All regressions include firm*individual fixed effects and gender-region specific time shocks – with the exception of Column 2 that controls for gender-1-digit SIC specific time shocks. SIC information is missing for 0.0002 percent of observations. Heteroskedasticity- robust standard errors clustered at firm level in parentheses. The p-value at the bottom of the table refers to the t-test on the sum of the two reported coefficients, corresponding to the effect of the policy on female employees. The pre-policy mean represents the mean of the outcome variable for the treated group between 2013 and 2017.

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

Table A9: Impact on log hourly pay - BSD firm size

	Main spec (1)	BSD conteporaneous firm size (2)
Treated firm*post	-0.029***	-0.025***
	(0.009)	(0.009)
Treated firm*post*fem	0.030**	0.027**
	(0.013)	(0.014)
Observations	35,092	34,925
Adjusted R^2	0.894	0.894
P-value Women Coeff	0.909	0.807
Men's pre-policy mean	15.94	15.94
Women's pre-policy mean	13.36	13.36

Source: ASHE, 2013–2021.

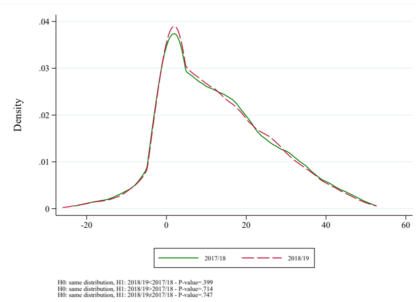
Notes: This table compares the main results on log hourly pay with the estimates from a specification where information on firms' numbers of employees is obtained from BSD contemporaneous firm size when it is missing in ASHE. The estimation sample comprises men and women working in firms that have between 200 and 300 employees. Both regressions include firm*individual fixed effects and gender-region specific time shocks. A treated firm is defined as having at least 250 employees in 2015. The post dummy is equal to one from 2018 onward. Heteroskedasticity-robust standard errors clustered at firm level in parentheses. The p-value at the bottom of the table refers to the t-test on the sum of the two reported coefficients, corresponding to the effect of the policy on female employees. The pre-policy mean represents the mean of the outcome variable for the treated group between 2013 and 2017.

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

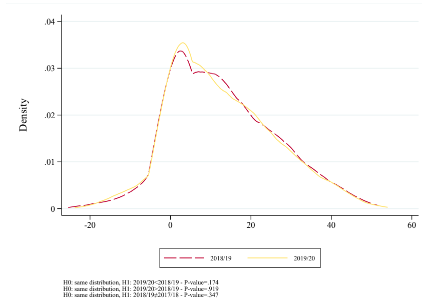
B Further information on mechanisms

B.1 Performance comparisons

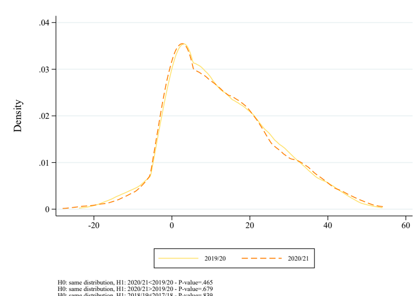
Figure B1: Gender pay gap distributions - year-on-year comparisons



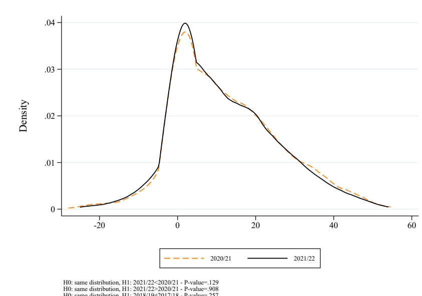
(A) 2017/18 vs. 2018/19



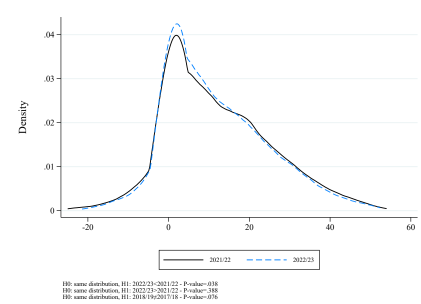
(B) 2018/19 vs. 2019/20



(C) 2019/20 vs. 2020/21



(D) 2020/21 vs. 2021/22

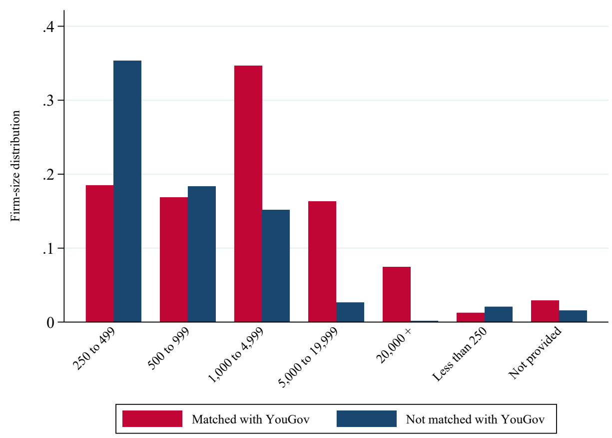


(E) 2021/22 vs. 2022/23

Source: UK Government Equalities Office, 2018-2023.
Notes: These graphs present year-on-year comparisons of the gender pay gap distribution. The data are drawn from the Gender Pay Gap Reporting website. Outliers (bottom and top 1 percent) are excluded from the graphs. The p-values reported at the bottom of each figure refer to the three hypothesis tested in the Kolmogorov-Smirnov test. The first test compares the null hypothesis that the two distributions are the same, relative to the alternative hypothesis that the distribution for year $t+1$ has *smaller* values than the t distribution. The second test compares the null hypothesis that the two distributions are the same, relative to the alternative hypothesis that the distribution for year $t+1$ has *larger* values than the t distribution. The third test compares the null hypothesis that the two distributions are the same, relative to the alternative hypothesis that the two distributions are different.

B.2 YouGov data and firms' reputation

Figure B2: YouGov vs. GPG sample firm-size distribution



Source: UK Government Equality Office, YouGov 2018-2019.
Note: These figures compare the firm-size distribution among GPG firms that match or not with YouGov.

Table B1: Gender equality performance and presence in YouGov

	2017/18				2018/19			
	Entire sample	Matched with YouGov		P-value difference	Entire sample	Matched with YouGov		P-value difference
		No	Yes			Yes	No	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender pay gap (%)	11.79	11.73	12.92	0.09	11.88	11.85	12.37	0.46
	(15.84)	(15.94)	(13.80)		(15.51)	(15.61)	(13.38)	
Observations	10,557	10,017	540		10,812	10,285	527	

Source: UK Government Equalities Office, YouGov 2018–2019.

Notes: This table explores potential selection patterns of GPG firms matched with YouGov. Column 1 (5) reports the gender median hourly pay gap for all GPG firms in 2017/2018 (2018/2019); Column 2 (6) refers to firms that we do not find in YouGov; Column 3 (7) refers to firms matched with YouGov; Column 4 (8) reports the p-value of the difference in the sample means of these two groups.

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

B.3 Firms' response to public scrutiny

To study firms' response to public scrutiny, we compare the impact of the policy on the gender pay gap across firms that are more or less likely to be exposed to public scrutiny. We proxy firms' exposure to public scrutiny by firms' pre-policy investment in advertising. Our hypothesis is that the public audience will be more familiar with businesses that spend more in advertising. In turn, these firms may also be more scrutinized by the public audience.

To retrieve information on firms' advertising expenditure in the pre-policy period, we used the Annual Business Survey. The Annual Business Survey (ABS hereafter) is an annual survey of businesses covering the production, construction, distribution, and service industries, which represent about two-thirds of the UK economy in terms of gross value added. Among other variables, ABS provides data on advertising costs and turnover. Importantly, ABS reports data at the establishment level, so for each firm and year, we first sum advertising costs and turnover at the firm level. Next, for each year and firm, we constructed an advertising-to-sales ratio, as the ratio between advertising costs and turnover, and computed the average ratio for each firm between 2013 and 2017.[A.1] Third, we matched ABS and ASHE, using the common anonymized firm identifier, and found 78 percent of firms included in the estimation sample. We then excluded firms in the top 1 percent of the distribution of the advertising-to-sales ratio (these are firms that spend more than 80 percent of their sales in advertising). Finally, we rank ASHE firms based on their average pre-policy advertising-to-sales ratio and grouped employers with below- and above-median advertising-to-sales ratios.

[A.1] Note that, because ABS is a survey, only a representative sample of firms is interviewed every year: when considering the pre-policy years 2013 to 2017, we found that 90 percent of firms with 200 employees or more are present at least 3 years in the survey, and 70 percent of them are present every year.

C Firms' hiring practices and gender equality

In this section, we explore whether firms' gender equality performance correlates with their hiring practices. A growing number of papers document that a factor contributing to the persistence of the gender pay gap is the so-called gender ask gap, whereby women shy away from wage bargaining or propose a lower ask salary when stating how much they want to make in their next job (Babcock et al. 2003, Hall and Krueger 2012, Leibbrandt and List 2015, Card et al. 2016, Roussille 2020, Biasi and Sarsons 2022). Upfront wage information in the recruitment process may help address this gap by reducing the room for wage bargaining. Consistent with this hypothesis, Flinn and Mullins (2021) show that in a labor market with heterogeneous wage settings, where both wage bargaining and wage posting initially coexist, mandating wage posting reduces the gender pay gap by 6 percent.[A.2] It is also interesting to note that the European Commission has recently issued a directive that nudges firms to post wage information in job listings, as part of a series of pay transparency measures aimed at improving gender equality in the labour market.[A.3] To explore how wage posting correlates with firms' gender equality performance, we combine the GPG data on firms' equality indicators with Lightcast job vacancy data.

Lightcast, previously known as Burning Glass Technologies, scrapes online job ads from company websites and job boards. UK data are available from 2012 and cover more than 50 million (de-duplicated) individual job vacancies collected from a wide range of online job listing sites. While the data set only includes online advertisements, and hence misses vacancies not posted online (e.g. those advertised informally and internal vacancies), it includes a rich set of information that is especially useful for our analysis. First, each observation includes the text of the job advertisement. Second, more than 95 percent of vacancies have an occupational SOC identifier. Third, around one third of the vacancies, or 20 million observations, include the name of the employer. As this is the only variable that can facilitate the merging of Lightcast data with other firm-level data, we focus on the restricted sample with non-missing employer names. We also exclude vacancies posted prior to 2014, as Lightcast expressed concerned over the quality of the data at the beginning of the sample (Adams-Prassl et al. 2023).

To study how firms' wage-posting decision correlates with their gender equality performance, we extract wages offered from the job-ad text using natural language processing. In particular, to identify wages in the text, we use a series of targeted regular expressions that pick up phrases such as "30-35k per annum" and "20,000/year". A series of validation exercises conducted by research assistants show that we correctly classify the presence of wages in 98 percent of cases. The residual 2 percent are false negatives, meaning that our code indicates that there is no wage posted when there actually is one. In addition to these validation exercises, we also exclude vacancies in the bottom 5 percent and top 1 percent of the distribution of posted salaries. We then measure wage posting using a dummy variable equal to one if the vacancy contains wage information, either in the form of a wage interval or a point offer.

As a last step, we match GPG firms' gender equality indicators with Lightcast data using the name-matching strategy explained below. We retain only employers with a match score of one and non-missing SIC and SOC information, for a total of 7,126 GPG firms. Table C1 explores selection

[A.2] Though, importantly, directed search models and related empirical evidence show that wage posting may increase competition for a job (Banfi and Villena-Roldan 2019, Marinescu and Wolthoff 2020, Wright et al. 2021, Belot et al. 2022).

[A.3] The European Commission directive is available at https://ec.europa.eu/commission/presscorner/detail/en/ip_22_7739.

patterns of the matched sample. While GPG firms matched with Lightcast have, on average, a larger and statistically different gender pay gap than firms that do not match with Lightcast, the percentage of women in the top quartile of the firm wage distribution is not statistically different across the two groups.

The bar graph in Figure C1 reports the correlation between GPG firms' average percentage of vacancies posting wage information between 2014 and 2021 and, respectively, the average percentage of women in the top quartile of the firm wage distribution (blue bar), and the average gender pay gap (red bar) between 2018 and 2021. When computing these correlations, we control for firms' 5-digit SIC codes, the average number of vacancies a firm posts per year, and the occupational composition of vacancies; we also cluster the standard errors at the 5-digit SIC level. While in Lightcast data we find that only 50 percent of firms post wage information, the graph shows that firms that are more likely to do so also tend to have a larger percentage of women at the top of the firm wage distribution and a lower gender pay gap. Although these are only correlations, they are consistent with the hypothesis that wage posting may help address the gender ask gap.

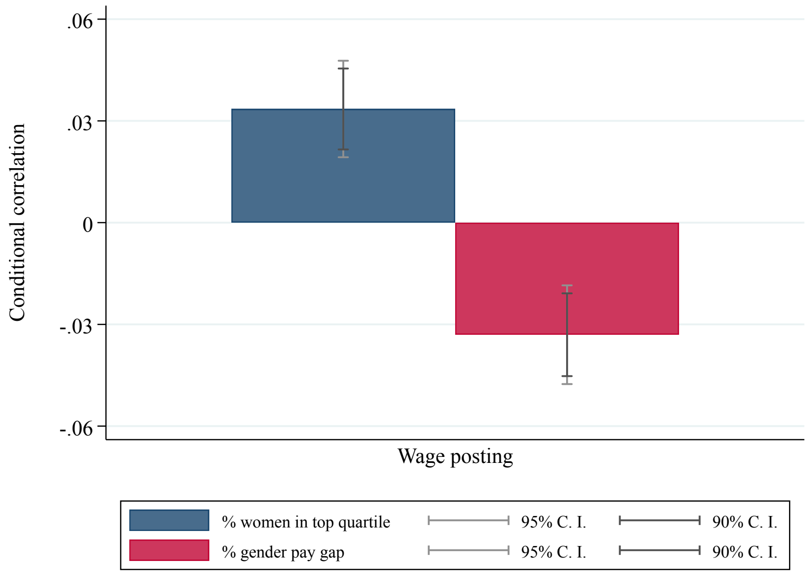
C.1 Name matching algorithm

We merge two different firm-level data sets, A and B, through the only common identifier available: firm name. We first collapse all firm names in each data set down to a unique set of firm names using standard text cleaning procedures; this includes dropping any exact duplicates. We then use firm names from one of the datasets, A, to define a vector space using all character-level 1–4-grams (with a maximum of 30,000 features) to create a matrix with dimensions number of entries in A times number of text features. This is achieved using Python's scikit-learn's (Pedregosa et al., 2011) TF-IDF Vectorizer, so that frequently appearing 1–4 character grams are down-weighted. As the final stage of preparation for matching, the cleaned firm names in B are expressed in the vector space defined by the cleaned firm names from A.

To perform the matching, we use cosine similarity. Note that this involves taking the inner vector product of every firm name in A with every firm name in B so is computationally intensive. To facilitate this, we use the sparse_dot_topn package, developed by ING Bank, to perform parallel computation of the closest matches across A and B.

The result is an array of scores of the firm name matches between A and B that we are then able to use at different thresholds according to how close a match we prefer, with unity reflecting a perfect match in the vector space, and 0 reflecting two firm names that are entirely orthogonal in the vector space.

Figure C1: Wage posting and equality indicators - conditional correlations



Source: Lightcast 2014–2021. GPG 2018–2021.
Note: The bar graph reports estimated coefficients from regressions of gender equality indicators (averaged across 2017/18 and 2020/21) on the average percentage of vacancies posting wage information over the period 2014–2021, the occupational composition of firms' vacancies, firms' average annual number of vacancies, and 5-digit SIC fixed effects. The graph also displays 90 and 95 percent confidence intervals associated with heteroskedasticity-robust standard errors. The sample includes firms publishing gender equality indicators between 2018 and 2021, matched with Lightcast with a match score of 1 (See Appendix Section C.1 for a description of the name-matching procedure), and non-missing SIC and SOC codes. Vacancies with salary outliers(bottom 5 and top 1 percent) are also excluded from this analysis. N. observations = 7,126.

Table C1: Gender equality performance and presence in Lightcast

	Entire sample	Matched with Lightcast No	Yes	P-value
	(1)	(2)	(3)	(4)
Gender pay gap (%)	12.04	12.03	12.05	0.94
	(15.11)	(15.46)	(14.77)	
Women in top quartile (%)	40.08	39.96	40.20	0.57
	(24.01)	(24.18)	(23.86)	
Observations	13,849	6,732	7,117	

Source: Lightcast 2014-2021, GEO 2018–2023.

Notes: This table explores potential selection patterns of GPG firms matched with Lightcast. Column 1 reports the average gender median hourly pay gap and percentage of women in the top quartile of the firm wage distribution for GPG firms publishing equality indicators between 2018 and 2021; Columns 2 and 3 compare equality indicators across GPG firms that match with Lightcast or not; Column 4 reports the p-value of the difference in the sample means of these two groups.

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