

Department of Economics, University of Warwick
Monash Business School, Monash University

as part of
Monash Warwick Alliance

**Is French Gender Pay Gap Transparency Legislation
Effective?**

Mia Vorster

Warwick-Monash Economics Student Papers

March 2026

No: 2026/93

The Warwick Monash Economics Student Papers (WM-ESP) gather the best Undergraduate and Masters dissertations by Economics students from the University of Warwick and Monash University. This bi-annual paper series showcases research undertaken by our students on a varied range of topics. Papers range in length from 5,000 to 8,000 words depending on whether the student is an undergraduate or postgraduate, and the university they attend. The papers included in the series are carefully selected based on their quality and originality. WM-ESP aims to disseminate research in Economics as well as acknowledge the students for their exemplary work, contributing to the research environment in both departments.

Recommended citation: Vorster, M. (2026) Is French Gender Pay Gap Transparency Legislation Effective? *Warwick Monash Economics Student Papers* 2026/93

WM-ESP Editorial Board

Sascha O. Becker (University of Warwick)

Mark Crosby (Monash University)

Cecilia T. Lanata-Briones (University of Warwick)

Gordon Leslie (Monash University)

Thomas Martin (University of Warwick)

Vinod Mishra (Monash University)

Jeremy Smith (University of Warwick)

Natalia Zinovyeva (University of Warwick)

If you want to get in touch with the Editorial Board, please email w-mesp@warwick.ac.uk

Is French Gender Pay Gap Transparency Legislation Effective?

Mia Vorster *

September 2025

Abstract

This paper examines the impact of France's 2018 gender pay gap transparency law which implements a 100 point gender equity index and two corrective measure thresholds (firms below 85 must publish an action plan; those below 75 for three years are fined) for firms with more than 50 employees. I use firm-level reports with a regression discontinuity design and event-time study to assess the effectiveness of the thresholds in changing firm behaviour, finding evidence that firms improve their gender equity in order to avoid fines below the 75 threshold. I use an individual-level administrative dataset which I transform into a pseudo-panel and a difference-in-difference strategy around the 50 employee threshold to assess the overall impact of the policy, finding a 3.7 percentage point increase in women's hourly wages relative to men in treated firms. This effect is primary driven by a fall in male wages.

Keywords: Gender Pay Gap, Pseudo-Panel, Wage Transparency, Public Policy, Gender Pay Gap Reporting

JEL Classifications: J16, J78, K38

*mia.vorster1@outlook.com

Note: the Appendix is found at <https://www.dropbox.com/scl/fo/qinwqjrctnyvz53jog7iy/A0WE8EhqBMo0Bk1dKnSQAmA?rlkey=mrqbepta96m3az5e7h5m49ia1&st=y7fc3mfh&dl=0>

1 Introduction

Pay discrimination on the grounds of gender is prohibited by the French constitution (Schäfer and Gottschall, 2015), yet it both persists and is among the highest in Europe (World Economic Forum, 2024): while the EU has an average pay gap of 13%, in France it is 15.8%, putting France 18th in the EU (European Commission, 2022). The Global Gender Gap Report 2025 estimates global gender parity to be over a century away (World Economic Forum, 2025). People underestimate gender pay gaps (GPGs): survey participants estimated an average within-firm pay gap of 6.7%, almost half of what it really is, dampening pressure on firms to act and helping gaps persist (Blundell, 2021).

Pay gaps remain a policy concern. The European Commission issued recommendations in 2014 and 2017, and a binding Pay Transparency Directive in force in 2026. In 2018, France passed the Professional Futures Act—a policy requiring firms with 50+ employees to publish a 100-point annual gender equity index—following similar legislation from countries such as the United Kingdom, Denmark, Austria, and Germany. Unlike these transparency-only directives—where reporting is required, but no remediation *must* occur—in France, firms must implement improvement plans if their score is below 85, and those below 75 for three consecutive years face fines. These create compliance incentives beyond reputational ones. While France has recently announced that it is moving away from its current index (Bissuel and Métais, 2025), but a study of its novel penalty system can help inform the design of new mechanisms.

This paper asks whether France’s policy reduces pay gaps overall, and whether the thresholds alter behaviour amongst the worst performers. I combine three empirical strategies. Using the firm-level indexes, a regression discontinuity design around the thresholds examines whether index growth differs immediately above versus below them; firms just below 75 points grow 1.6 index points per year faster than those just above, consistent with penalty avoidance behaviour. An event study of threshold crossers shows that crossings are precipitated by large one-year negative shocks, not gradual drifts, and that firms falling below 75 experience larger negative shocks but faster catch-up than 85-crossers, again consistent with fine-avoidance. Finally, a difference-in-differences design using employee-level administrative data indicates a policy effect of roughly 3.7 percentage point increase in women’s hourly earnings relative to men’s amongst affected firms, driven by a fall in male wages.²

In Section 2, I review the literature that exists on this topic, before describing the legislation in detail in Section 3. In Sections 4 and 5 I walk through my data and methodology, before discussing results in Section 6.

2 Literature Review

France’s performance in global gender equality rankings presents a mixed picture. From the Global Gender Gap Report 2024 (World Economic Forum, 2024), France ranks 22nd overall and 11th within the G20, yet is only 100th for wage equality. Women achieve higher education, with France first in educational attainment, yet remain under-represented in science and

²Other literature, e.g. Blundell et al. (2025), suggests that this is a reduction in male wage growth, rather than wages *per se*, but this is not within the scope of this paper.

technology fields, accounting for only 39.2% of graduates. Hofstede’s *Power Distance* score for France shows that cultural factors may affect the persistence of these gaps: France scores 68, suggesting greater acceptance of hierarchy and inequality than the UK or Germany, both at 35, reducing societal pressure to close gender wage gaps.

One of the major factors behind women’s lack of wage equality is heterogeneities in labour force participation, which is 56% for men versus 49.1% for women (International Labour Organization, 2025). Women are overrepresented in part-time work: 33% of employed women work part-time, compared to only 7% of men, making women 80% of all part-time employees, with 45% of part-time women having three or more children (Caudill et al., 2023). Occupational segregation is pronounced: women comprise 18.3% the industrial sector, but 81.7% of the service/agricultural sectors where wages are lower (International Labour Organization, 2009). Segregation accounts for approximately half of the explained gender pay gap, with the remainder due to individual worker characteristics (21%) and unexplained discrimination (29%) (Caudill et al., 2023). Discrimination compounds the effects of segregation, with some women working in low-paid jobs due to hiring barriers for higher-paying sectors rather than inadequate skills (Gobillon et al., 2015). Additionally, women cluster to low-paying firms when job choices prioritise family-friendly conditions rather than salaries (Groshen, 1991); heterogeneities in job characteristic weighting accounts for as much as 30% of the pay gap (Blau and Kahn, 2000) with societal norms explaining many of these differences. Nergaard (2002) find that women are pushed to work in the public sector as they are viewed as being more virtuous and caring. Bertrand et al. (2015) reports that women willingly give up income in order to not earn more than their husbands, a pattern explained by gender identity norms.

In 2017, President Emmanuel Macron declared gender equality the “great national cause” (Ministry for Europe and Foreign Affairs, 2021), with the wage gap in 2023 standing at 22.2% overall and 3.8% for employees in the same establishment and job (INSEE, 2025). Before 2018, the Labour Code (introduced 1973) mandated firms with ≥ 11 employees to hold annual consultations with their Social and Economic Committee³ on gender equality (Article L2312-26), and to conduct regular negotiations with labour unions on the subject, with non-compliance risking payroll fines of up to 1%. The 2011 *Coppé-Zimmerman Law* requires executive and supervisory boards of firms with more than 500 employees to comprise at least 40% of each gender, potentially reducing the upper-level gender pay gap; in the US, the largest pay gap is amongst the highest earners (Blau and Kahn, 2017). In 2018, the Professional Futures Act was adopted, requiring firms over 50 employees to annually publish a composite gender equity index, scored out of 100. It comprises five subindices, i) the pay gap, ii) the gap in salary increases, iii) the promotion gap, iv) raises following returns from maternity leave, and v) the gender balance in the top 10 earners. Two thresholds exist: firms scoring below 85 are mandated to publish gender equity action plans, while firms below 75 must correct this within three years or face annual payroll fines up to 1%. Implementation was staggered by firm size. Section 3 goes into greater detail.

There is a broader literature examining the effects of pay transparency. In Canada, Baker et al. (2023) find that university pay transparency reduces the pay gap amongst staff. Mas (2014), studying 2010 Californian pay transparency legislation, reports evidence of wage com-

³A Social and Economic Committee (CSE) is an employee representative body inside a company, elected by its employees. It covers issues such as health and safety, working conditions, and social dialogue between employer and employees.

pression amongst highly-paid managers—which has implications for the subindex measuring the top ten earners’ gender distribution. Galanakis and Gosling (2024) finds that the gender composition of corporate boards affects wages, the reported gender pay gap, and the interest in mandatory pay gap reporting. Female directors act as a signal of commitment to equitable compensation, affecting workplace culture and influencing worker preferences. Cullen and Perez-Truglia (2018) show that workers know they earn less than their peers reduce their effort, suggesting a mechanism where, with mandatory GPG transparency, firms may narrow pay gaps to restore productivity. Similarly, Cullen and Pakzad-Hurson (2023) find that higher transparency results in more equal but lower wages. Nonetheless, these findings have limited direct applicability to my study: these legislations are *individual-level* transparency, i.e. the revealing of certain individual’s wages directly, rather than mandating the creation of aggregate firm-level statistics.

More directly relevant, Bennedsen et al. (2022) find that Denmark’s wage transparency legislation increased female earnings by 2 percentage points relative to men. Blundell (2021) found that the equivalent British policy narrowed the gap by 1.6 percentage points. These papers, with Blundell et al. (2025), identify the mechanism as a reduction in male wages—specifically, a reduction in the male salary growth rate—rather than an increase in female wages, a pattern consistent with findings in Cullen and Pakzad-Hurson (2023). Blundell (2021) also shows that workers forgo income to avoid firms with large pay gaps, explaining why poorly performing may narrow gaps following transparency mandates, while Bennedsen et al. (2022) finds a negative effect on firm productivity. Jones et al. (2023) leverage an unexpected COVID-19-induced suspension of British reporting requirements to show that firms which report had 6 percentage point lower gender pay gaps than non-reporters.⁴ Gulyas et al. (2023) and Böheim and Gust (2021) study the Austrian transparency law, finding no significant impact, likely because transparency only extends to employees of that firm, not the wider public, reducing reputational pressure. The European Commission issued non-binding recommendations on the gender pay gap in 2014 and 2017, and has adopted the Pay Transparency Directive for 2026, requiring all member states to implement legislation aimed at closing pay gaps. Iceland’s 2018 legislation requires firms to demonstrate equal pay, shifting the burden of proof to employers. Finland, meanwhile, awards the International Gender Equality Prize biennially to organisations that have made significant contributions to gender equality, carrying a reward of €300,000.

While prior studies have examined national gender pay gap legislations, this paper is the first to analyse the French Professional Futures Act. Unlike earlier cases, this legislation incorporates explicit penalty thresholds for poorly performing firms, enabling an assessment of whether financial sanctions change firm behaviour. I contribute to the literature firstly by a causal evaluation of the legislation’s impact on the gender pay gap, and secondly by isolating the behavioural response to crossing penalty thresholds, a feature absent from previously studied policies. In the next section, I walk through the specifics of the French legislation, before going on to discuss my data and methodology.

⁴The unexpected suspension was late enough that many firms had already reported, so the paper compares 2021/22 pay gaps of firms which ‘accidentally’ reported in 2020/21 and those that chose not to.

3 Policy Overview

The Professional Futures Act (*Loi pour la liberté de choisir son avenir professionnelle Law n°2018-771*) was enacted on the 5th of September, 2018, coming into force on the 1st of January, 2019. Firms with more than 50 employees must report a gender equity index, scored out of 100⁵, which is comprised of disaggregated statistics with have unequal weights:

1. Wage equality: 40 points
2. Gap of individual salary increases: 20 points
3. Promotion gap: 15 points
4. Percentage of employees whose salary was increased upon their return from maternity leave: 15 points
5. Gender balance among the ten highest paid jobs: 10 points

For firms under 250 employees, the salary increase and promotion gaps are combined to create a 35-point subindex.

Policy implementation was staggered by firm size:

- ≥ 1000 employees: first publication by 1st March, 2019
- 250–999 employees: 1st September, 2019
- 50–249 employees: 1st March, 2020

Firms can choose any consecutive 12-month reference period to calculate their index. For example, a firm with 50 to 249 employees publishing its first index by 1st March 2020 could have used data from as early as March 2018 to February 2019, or as late as March 2019 to February 2020. In practice most firms likely use the 2019 calendar year.

Subindex scores are arbitrarily defined, rather than a full score being total equality and 0 being total inequality: for instance, to score 20 for their salary gap, a firm must have a gap $\leq 2\%$, whereas 0 points is a gap $> 10\%$. Appendix A contains a complete breakdown of subindex scoring.

Firms report their overall index and a subindex breakdown on their own website,⁶ and to the Social and Economic Committee and the Labour Inspectorate, which publishes them via the Ministry of Labour and Employment’s *Egapro* tool, enabling users to easily find the scores of any reporting firm. The Ministry also publishes a combined dataset containing the indexes, which I use for part the analysis within this paper.

The French index falls between the British legislation and the Austrian or Danish legislations for its transparency impact. The British index is freely available on the Government’s website, but it provides more detailed information, reporting both median and mean wage differences,

⁵Where 100 is the best; 0 the worst.

⁶If a firm does not have a website, they must inform employees via alternative means.

and the by-pay quartile gender composition of the company. The British index applies to all firms over 250 employees, but due to structural economic differences this amounts to roughly the same number of employees in both economies.⁷ By contrast to the British and French indices, the Austrian and Danish ones are far less transparent: individuals can only access the results of the firm at which they are employed, and employees who disclose their firm's data face fines. Hence, French firms will experience greater public pressure to improve their gender equity practices.

The French legislation may hold greater power than the British index as there are correction measures and progression objective requirements for low-scoring firms. Firms that have obtained an index of less than 75 points must define 'adequate and relevant corrective measures' in order to improve within three years, or otherwise face a potential fine of up to 1% of their payroll annually until correction. Any firms scoring less than 85 points must set 'progress objectives' for each of the subindices whose maximum score has not been reached. Correction measures are defined in the context of compulsory negotiation between firms and employees on professional equality, or otherwise by unilateral decision by the employer after consultation with the Social and Economic Committee (CSE). They must be published on the company's website or otherwise communicated to employees, and are filed with Regional Directorate of Economy, Employment, Labour, and Solidarity (Dreets).

The index has had criticisms levied against it for its composition (Connolly, 2024), especially by trade unions and, recently, the High Council for Equality between Women and Men (HCE), an independent advisory body (Chetrit, 2024). The two salary increase indicators and the promotion indicator are agnostic about the *size* of the increases, rather indicating a binary 'percentage of employees who did/did not receive a wage'. For example, a firm could give women raises exactly as often as men, scoring the maximum 15 points, and the index would not care if female raises were half the size of male raises. Firm index growth over time may be partially a result of better data manipulation strategies rather than representing actual improvement. Employees receiving raises after returning from maternity leave is already a legal requirement, thus making firms which do not score 15 in breach of the law. Because the measure is presented as a composite index, scores alone do not convey the underlying statistic. Interpreting it requires consulting the breakdown of subindex scores⁸ or conducting a separate search for the detailed data. This extra step may dissuade immediate understanding by individuals, limiting the potential impact of the index on public awareness and firm behaviour. Finally, the limited granularity in the index hampers detailed analysis of the firm-level data. While the pay gap indicator allows any integer score from 0 to 40, all other subindices are scored in increments of five points, except for the post-maternity leave raise indicator, which is effectively binary, awarding either 15 points or none, for full or less-than-full compliance respectively.

⁷56% of British employees work in firms larger than 250 employees (Office for National Statistics, 2022); 46% of French employees work in firms over 50—a statistic generated using data from my individual-level data source, *Base tous salariés*.

⁸This involves navigating to a separate webpage whose link is in the small print at the bottom of the Egapro website. It is not difficult to find for a French-language speaker.

4 Data

4.1 Firm-level data

To assess the effectiveness of the 85- and 75-point thresholds in changing the behaviour of affected firms, I use data from the *Egapro Index de l'égalité professionnelle* published by the French Ministry of Labour and Employment. The dataset contains the submissions of all firms legally required to report under the Professional Futures Act, ensuring complete coverage of the reporting population. Firms with at least 50 employees are obliged to submit their results annually, with non-compliance subject to sanctions. The dataset covers the period 2018-2024, corresponding to the reporting years 2019-2025, reflecting the legislation's staggered implementation.

For each firm, Egapro reports the overall composite index (0 to 100 points), and the five underlying subindices. Administrative variables include the firm's department and region, their NAF code (industry classification) and employee headcount category⁹. While the gender equality subindex, scored out of 40, can be scored as any integer, the salary raise, promotion, and gender balance of the ten highest-paid employees are reported in five-point increments, or, in the case of the post-maternity raise calculator, as a binary score (0 or 15 points). Subindices may be reported as 'non-calculable' if the legal thresholds for calculation are not met (e.g. fewer than the minimum number of male and female employees in a category).

In this analysis, I focus on the combined index to evaluate the thresholds, using firm-size, region, and a simplified industry code as controls in my regressions. While the self-reported nature of the data introduces worries for its accuracy, the full-population coverage makes Egapro a robust source for studying the policy's effects on firm behaviour. I consider only metropolitan France, dropping any company which is placed in areas such as French Guiana. The dataset includes foreign-headquartered companies that operate in France and are subject to the same reporting obligations, but these firms have no associated French regional code, so are dropped in specifications that include region fixed effects. Firms whose combined index scores are non-calculable are likewise unable to be analysed.

4.2 Employee-level data

To assess whether the policy reduced the overall wage gap in affected companies, I use the *Base tous salariés*, an administrative dataset produced by INSEE from mandatory employer declarations (*Déclarations Sociales Nominatives*), covering approximately one twelfth of the French workforce annually. There are two versions of the data: one at the individual level individuals (where individuals may hold multiple jobs) and one at the job level. I use the latter, and restrict my analysis to the full-time workforce to minimise the possibility of including multiple jobs held by the same person.¹⁰ The data is released as separate annual files, though with consistent methodology; I download the datasets for each year from 2015 to 2022 individually and combine them into a single dataset for analysis. There are roughly three million observations per year before restrictions.¹¹

⁹50-249, 250-999, 1000+

¹⁰As a robustness check, I perform regressions on the full-time workforce.

¹¹After dropping non-full time jobs, this became roughly 1.9 million.

The main variables of interest are the exact salary of each employee, their their hours worked, legal sex, and their firm size in tranches. These will construct my dependent variable and the triple interaction. Salaries are deflated using INSEE’s base-2015 consumer price index, and I construct an hourly pay variable and use its natural logarithm in the main specifications. Control variables include the socio-professional category of the employee (PCS code, with separate 29 categories), the industry of the employing firm (the NAF Code, using the A17 specification with 17 industry-types), and the individual’s department of residence. I construct a private sector dummy variable, using data on the public-private status of an employee’s firm.

The dataset is anonymised: it contains no individual identifiers, so individuals cannot be tracked over time, making it a repeated cross-section rather than a true panel. In the full version of the dataset, individuals can be followed over time, but the anonymised version available for this analysis removes all identifiers, making such tracking impossible. It also contains no firm identifiers, preventing the use worker-firm fixed effects or the calculate employer-level wage gaps. To enable longitudinal analysis, I construct a pseudo-panel following the methodology of Guillerm (2015, 2017) and Meng et al. (2014), described in detail in Section 5.2.

Again, I restrict my data to metropolitan France. Following Blundell (2021), I include only individuals aged 19 to 55, excluding those aged 18 and under as well as those aged over 55. This reduces the influence of school-to-work transitions and retirement-related labour market exit on measured wage gaps. Retirement age is 62 in France, but individuals begin to leave the labour market before reaching this age, meaning that the older age workforce is increasingly subject to selection and so the population becomes less representative. Selection is a wider issue in the labour force: women who experience significant wage gaps are more likely to leave it, upward-biasing women’s wages and downward-biasing the wage gap. There is a chance that a pay gap transparency policy may attenuate the statistical wage gap, precisely because it helps to close it: that is, women who would otherwise leave because of large gaps now choose to stay.

5 Methodology

5.1 Firm-level analysis

My firm-level analysis examines the causal effects of the penalty thresholds in the *Index de l’égalité professionnelle* on firm behaviour. I focus on the two cutoffs: the 75-point threshold, below which firms must improve within three years or face fines, and the 85-point threshold, which mandates the creation of action plans. Using firm-level index scores from Egapro dataset, I track the same firms over time, enabling direct examination of behaviour changes around these discrete policy thresholds.

I employ two complementary approaches to assess the causal effects of the thresholds on firm behaviour. Firstly, a regression discontinuity design (RDD) to estimate the local, immediate impact of crossing a threshold on the subsequent index growth. Secondly, an event study traces the dynamic path of index scores before and after a threshold crossing, providing insight into both pre-treatment trends and the persistence of post-treatment effects. The RDD shows the local effect at the cutoff, while the event study shows the timing and evolution of impacts,

making the two approaches jointly informative.

5.1.1 Regression discontinuity design

My outcome variable is the annual change in the composite index, and the running variable is the lagged index score. This estimates the local average treatment effect of being just above the threshold in period $t - 1$ on index growth in period t , compared to being just below it.

In the standard RDD framework, ‘treatment’ is defined as being above the threshold. In my context, however, policy mandates apply to firms scoring *below* the threshold. For consistency with conventional notation, I maintain the standard orientation, so regression coefficients are interpreted as the difference in outcomes for firms above the threshold relative to those below it.

Thus, my specification is

$$\begin{aligned} \Delta \text{Index}_{it} = & \alpha + \tau [\text{Index}_{i,t-1} \geq 75] + \beta_1 (\text{Index}_{i,t-1} - 75) \\ & + \beta [\text{Index}_{i,t-1} \geq 75] (\text{Index}_{i,t-1} - 75) + \varepsilon_{it} \end{aligned} \tag{1}$$

where ΔIndex_{it} is the outcome variable, index growth from $t - 1$ to t , $\text{Index}_{i,t-1}$ is the running variable, and $\text{Index}_{i,t-1} \geq 75$ is the cutoff.

Because the composite index is calculated from coarse subindices—most in multiples of five—the total score distribution creates bunching, especially around the top of the index distribution.¹² A Cattaneo et al. (2018) test for manipulation in the running variable confirms significant discontinuities at 85, consistent with bunching, but not at 75. Therefore, I restrict the RDD analysis to the 75 point threshold.

I implement a sharp RDD, as the policy takes effect the moment a firm goes below the threshold, using a triangular kernel and mean squared error-optimal bandwidth selection following Calonico et al. (2014). Models are estimated with and without controls for firm size, region, and simple industry fixed effects.

5.1.2 Event study

To complement the local RD estimates, I estimate a fixed effects event study model that tracks the trajectory of index changes relative to the first year a firm’s composite score falls below either threshold, using a three-year window on either side of the crossing. This hopes to answer the question ‘*how does a firm’s equality index evolve over time before and after that firm crosses a threshold?*’.

The dependent variable is the change in the index, and event-time dummies are defined relative to the first year a firm’s composite score falls below the threshold. My generalised specification is as follows

¹²For instance, a firm may score full, or near-full points on the wage gap score, but lose 5, 10, or 15 points from other subindices.

$$\begin{aligned} \Delta\text{Index}_{it} = & \alpha + \beta_3(\text{Lag } 3)_{it} + \beta_2(\text{Lag } 2)_{it} \\ & + \gamma_0(\text{Lead } 0)_{it} + \gamma_1(\text{Lead } 1)_{it} + \gamma_2(\text{Lead } 2)_{it} + \gamma_3(\text{Lead } 3)_{it} \\ & + X'_{it}\Gamma + \mu_i + \lambda_t + \varepsilon_{it} \end{aligned} \quad (2)$$

where ΔIndex_{it} is the change in a firm’s index score from $t-1$ to t . $J = 3$ lags and $K = 3$ leads are considered relative to the first threshold crossing, with $\text{Index} = 85$ in my first specification and $\text{Index} = 75$ in my second. In the combined specification, indicators for both thresholds are included in order to account for firms that cross both. Following convention, $t = -1$ is the reference period (Clarke and Tapia-Schythe, 2020): all coefficients will be relative to $t = -1$ levels. X' are the controls, and μ and λ are firm and year fixed effects, respectively, to absorb time-invariant firm characteristics and common shocks.

I control for ‘headroom’ (100 minus the lagged index) to capture the mechanical ceiling effect—firms nearer 100 have less scope to improve—as well as its root ($\sqrt{100 - L.\text{Index}}$) to account for diminishing returns to headroom—headroom in a 1-point change near the ceiling (e.g. index 95 to 96) matters more than the same change at a lower score (index 40 to 41). Including these controls isolate changes attributable to the policy itself.

The coefficients on the post-crossing leads (γ_k) estimate the average treatment effect on the treated (ATT) at each relative time period while lags (β_j) allow for inspection of pre-trends. These are plotted to display the trajectory of index changes before and after the crossing.

My main event study specification uses an unbalanced panel, so the set of firms observed varies across event-time periods: I include all available firm-year observations within the -3 to +3 period, even if they do not appear in all event years. A -3 to +3 balanced panel would require me to restrict my regression to only firms over 1000 employees (the only reporters in 2018) which crossed the threshold in 2021, which would introduce bias from year effects.

An unbalanced study may introduce compositional effects, where changes in firm composition over event time affects the results, so I include a model with a restricted subsample of firms observed in every year from $t = -2$ to $t = +2$, keeping $t = -1$ is the reference period, as a robustness check.

5.2 Employee-level triple difference model

My objective is to establish whether the gender equity policy helped to reduce the overall gender wage gap in affected companies. My data covers individuals aged 19-55 in the full-time workforce. I focus on firms in size tranches 20-99 employees, using 20-49 as the control group and 50-89 as the treated group; exact headcounts are not available in the anonymised data so an RDD is not possible. The anonymised *Base tous salariés* is a repeated cross-section, so does not allow individuals to be tracked over time; hence, I construct a pseudo-panel to mimic the properties of a true panel. This allows me to follow fixed cohorts over time, controlling for unobserved time-invariant heterogeneity, and reducing bias from year-to-year compositional changes.

Following the recommendations of Guillerm (2015, 2017) and Meng et al. (2014), I construct synthetic cohorts along time-invariant characteristics, meaning that individuals appearing in multiple years, even though untracked in the original dataset, will always reappear in the same

cohorts. These are defined on birth year, legal sex, and department of residence. These latter two vary, but do so infrequently enough to be treated as fixed, especially over the short time period analysed. In pseudo-panel analysis, cohorts are designed to group individuals with similar behaviours, so greater granularity through the definition of highly specific cohorts is preferable; however, there exists a trade-off where smaller cohorts increase the risk that empirical cohort means deviate from the true cohort means, introducing bias and reducing precision (Verbeek, 2008). Meng et al. (2014) drops all cohorts with fewer than 30 individuals, while Verbeek and Nijman (1992, 1993) recommend that cohorts should optimally contain at least 100 individuals. To increase cohort sizes, I group years of birth in two-year bands¹³. My final cohort construction creates 28,051 cohorts ($\approx 3,500$ per year) with a mean size of 163 individuals; 1,653 cohorts with fewer than 30 individuals are dropped.

Cohort means are calculated for continuous variables: log hourly pay for the primary specification, log annual salary for robustness checks, and log hours as a control for the latter.¹⁴ Categorical control variables—industry and job classification—are converted into factor variables and averaged to obtain cohort-level proportions, alongside a private sector dummy and a treatment dummy which are also proportionalised.

I estimate a triple difference specification, interacting treatment status—defined, after cohort-level dataset collapse, as the proportion of a cohort employed in a treated firm—with ‘post’ dummy variable, and a female cohort dummy. Post is defined as 2020 or later, as all treated firms are below 250 employees and thus became subject to the policy in that year. Cohort sizes vary substantially,¹⁵ so cohort observations are weighted by the square root of cohort size (Deaton, 1985); unweighted regressions are presented later as a robustness check.

Before estimation, I conduct a Hausman specification test comparing random effects and fixed effects. The null that random effects is consistent was strongly rejected,¹⁶ so I adopted fixed effects throughout to control for unobserved, time-invariant cohort characteristics. Standard errors are clustered at the cohort level to account for serial correlation and heteroskedasticity within cohorts over time.

The following cohort-level regression is estimated:

$$\begin{aligned} \log(wage_{ct}) = & \beta_0 + \beta_1 \text{Treated}_{ct} + \beta_2 \text{Post}_t + \beta_3 \text{Female}_c \\ & + \beta_4 (\text{Treated}_{ct} \times \text{Post}_t) + \beta_5 (\text{Treated}_{ct} \times \text{Female}_c) + \beta_6 (\text{Post}_t \times \text{Female}_c) \\ & + \gamma (\text{Treated}_{ct} \times \text{Post}_t \times \text{Female}_c) + \delta' X_{ct} + \alpha_c + \lambda_t + \varepsilon_{ct} \end{aligned} \tag{3}$$

where c indexes cohort, and t indexes years. X_{ct} is the vector of cohort-level controls, α_c denotes cohort fixed effects, and λ_t denotes year fixed effects. γ is the primary coefficient of interest, capturing the effect of the policy introduction on female wages.

There are limitations of this approach. I cannot track individuals through firms, so I cannot use firm-level fixed effects (or interact firm-level and cohort level fixed effects), potentially

¹³i.e. 1968-69, 1970-71, 1972-73, ...

¹⁴These are the averages of the log of individual’s earnings, etc., rather than the log of average earnings.

¹⁵Average cohort size of 163, with a standard deviation of 141. The smallest cohort has 30 individuals within it, the largest has 1,309.

¹⁶ $\chi^2(51) = 2901.63, p < 0.001$

Variable	20 - 49 employees		50 - 99 employees	
	Mean	Std. Dev.	Mean	Std. Dev.
Female	0.4742	0.4993	0.5222	0.4995
Age	36.3088	10.5696	36.6841	10.5798
Salary	€19,513.61	€22,918.78	€20,443.35	€25,237.99
Hours worked	1,064.58	746.97	1,079.85	737.11
Hourly salary	€17.47	€13.47	€17.76	€15.05

Table 1: Summary Statistics: Firms with 20 to 49 employees vs. 50 to 99.

leaving unexplained intra-firm variance. Additionally, treatment assignment is approximate at cohort level, being an intra-cohort proportion rather than a binary indicator.¹⁷

5.2.1 Common trends

The parallel trends assumption is key with a difference-in-difference strategy, requiring that, in the absence of the policy, the gap between male and female wages would have followed the same trend in treated and untreated employers. Analysis is restricted to firms around the 50-employee boundary, but, because I do not have granular employee counts, I use the two tranches around the threshold: firms sized 20 to 49 as the control, and 50 to 99 as the treatment. There are more individuals in the control (314,000 per year) than the treatment (231,000 per year), though this is before I collapse the dataset into cohorts: post-collapse, there is a proportion of both in all cohorts. I do not use a broader, and symmetrical, 0 to 100 employee window to avoid excessive heterogeneity—firms with, say, fewer than 10 employees are very different from those with 90.

Although firms larger than 50 employees are subject to additional obligations (e.g. funding workers councils), there are no other contemporaneous reforms around the 50-employee threshold that would differentially affect these firms during this period. While there is known bunching of firms at 49 employees to avoid these obligations (Garicano et al., 2016), the lack of precise firm headcounts prevents the use of a ‘donut hole’ strategy as in Blundell (2021).

Table 1 reports pre-policy summary statistics for the restricted sample. Firms with 50 to 99 employees have marginally higher age, salary, and number of hours worked, but using the latter two to compute hourly salary, they are broadly similar (1% higher in the larger firm-size). They also have a higher proportion of women.

To assess parallel trends, I estimate a generalised triple-difference event study by replacing the post-policy indicator with a full set of year dummies, interacted with treatment status and the female cohort indicator:

¹⁷Firm size cannot be used as part of the cohort definition: cohorts must be approximately time-invariant, and it implausible that individuals never, or even rarely, transfer between firms of, say, 40 employees and those of 60 employees.

$$\begin{aligned}
\log(\text{wage}_{ct}) = & \beta_0 + \beta_1 \text{Treated}_{ct} + \sum_{t \neq 2019} \beta_{2,t} \phi_t + \beta_3 \text{Female}_c \\
& + \sum_{t \neq 2019} \beta_{4,t} (\phi_t \times \text{Treated}_{ct}) + \sum_{t \neq 2019} \beta_{5,t} (\phi_t \times \text{Female}_c) \\
& + \beta_6 (\text{Treated}_{ct} \times \text{Female}_c) + \sum_{t \neq 2019} \gamma_t (\phi_t \times \text{Treated}_{ct} \times \text{Female}_c) \\
& + \delta' X_{ct} + \alpha_c + \varepsilon_{ct}
\end{aligned} \tag{4}$$

ϕ_t are time fixed effects, and the summation over $t \neq 2019$ means that coefficients are estimated for all years except the base year, so each $\beta_{2,t}$, $\beta_{4,t}$, $\beta_{5,t}$, and γ_t measures the difference relative to 2019. Vectors β_2 , β_3 , β_4 , β_5 and γ contain one coefficient for each non-base year. Visual inspection of γ_t for years prior to 2020 provides a direct check of the parallel trend assumption: pre-policy coefficients should not be statistically different to zero, while post-policy trends should reflect the policy's impact.

5.2.2 Timing and anticipation

For my main specifications, Post_t is set to equal 1 from 2020 onwards. The first reporting deadline for firms between 50 and 249 employees was in March 2020 and referred to pay in 2019. The legislation was passed in September 2018, giving firms around four months before the start of the 2019 pay year to adjust wage-setting. As *Base tous salariés* records total annual pay, anticipatory effects appearing in 2019 would reflect only part of the year under the new incentives, making it unlikely that the full policy effect would be captured before 2020. The event study presented above provides a check: given 2019 is my index year, pre-2019 coefficients being significantly negative would imply that firms were able to adjust their behaviour before the first reporting deadline.

5.2.3 Spillovers

The Stable Unit Treatment Value (SUTVA) assumption is a key identifying assumption. It mandates the policy not to have affected workers at companies in the control group. Blundell (2021), in his similar discussion, suggests three reasons to question the assumption.

Firstly, if the policy is successful and treated firms raise wages for their female workers, smaller firms may be pressured to match raises to remain labour market competitive. Given France's high labour market frictions (Cahuc, 2024), adjustments are unlikely to be immediate but could produce spillovers that attenuate the estimated treatment effect. Second, public awareness of GPGs could be affected by the policy's introduction, through government communications, media coverage, and workplace discussions. Firms may be pressure to lower the pay gap, but, again, this would downward-bias the policy effect on treated firms. Finally, smaller firms may adjust in anticipation of crossing the 50-employee threshold, expecting to grow beyond it in the near future. These are all cases of firms below the threshold improving their gender pay gaps, and thus are all examples of SUTVA violations attenuating the policy

Table 2: Regression discontinuity design

	(1) No Controls	(2) Controls
RD Estimate	-2.829*** (0.740)	-1.648*** (0.587)
Observations	84,362	83,915

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Outcome variable: equality index growth (period $t - 1$ to t). Running variable: Index score in $t - 1$. Cutoff: Index = 75. Column (1) has no controls, (2) has controls of region, firm size, industry, and year. Triangular kernel, local linear regression, with standard variables clustered by firm. The regression considers a ‘baseline’ below the threshold, versus a ‘treatment’ above the threshold.

effects. For SUTVA violations to upward-bias the policy effects, firms below the threshold would have to *increase* their gender pay gaps because of it, which seems implausible.

6 Results

6.1 Regression discontinuity design

Table 2 reports regression discontinuity estimates of the effect of crossing Index = 75 on subsequent index growth; if a firm is below this threshold for three years, they get a 1% payroll fine, a motivation for score increase. The outcome variable is a firm’s equality index growth from period $t - 1$ to t , while the running variable is the Index score in $t - 1$. Standard errors are clustered at the firm level. Column (1) presents the baseline specification, yielding an estimate of -2.829, significant at the 1% level. Column (2) introduces controls for region, firm size, industry, and year, reducing the magnitude to -1.648, still significant at the 1% level, indicating that firms just above the threshold experience lower index growth than those just below it, a pattern consistent with firms near the threshold taking corrective measures to avoid a penalty by moving above it.

Figure 3 (Appendix B) plots the local polynomial fits on either side of the 75-point cutoff. It shows a modest visual jump, indicating that much of the identifying variation stems from within-firm threshold crossing; when the data is pooled, within-firm discontinuities are averaged out and the scatter looks smooth. I therefore place more weight on the formal local RDD estimates.

Summary statistics reinforce this interpretation. Mean index growth (from $t - 1$ to t) among firms above the threshold in the *previous* period was -0.16, a decline. In contrast, firms above the threshold in the *current* period grew by an average of 2.03 points. This asymmetry is consistent with firms near but below the threshold increasing their equality scores to avoid a penalty, while those already above it have weaker incentives to maintain rapid improvements.

To assess robustness, I estimate alternate specifications, reported in Table 3. Columns (1) and (2) vary the bandwidth from the mean squared error-optimal value of 3.825 to 5 and 10, respectively. With a bandwidth of 5, the estimate remains significant, though a bandwidth of 10 reduces the magnitude and renders it statistically insignificant. Columns (3) and (4)

Table 3: RDD robustness tests

	(1)	(2)	(3)	(4)	(5)	(6)
	Bandwidth	Bandwidth	Poly 2	Poly 3	Cutoff	Cutoff
	5	10			73	77
RD Estimate	-1.386*** (0.468)	-0.444 (0.327)	-1.603*** (0.593)	-2.356*** (0.824)	0.402 (0.603)	0.230 (0.386)
Observations	83915	83915	83915	83915	83915	83915

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Outcome variable is equality index growth (period $t - 1$ to t), the running variable is Index score in $t - 1$. (1) and (2) vary the bandwidth to 5 and 10. (3) and (4) vary the order of the local polynomial to 2 and 3. (5) and (6) shift the cut-off to 73 and 77 as placebo tests. Triangular kernel, local linear regression, with standard variables clustered by firm. The regression considers a ‘baseline’ below the threshold, versus a ‘treatment’ above the threshold.

increase the order of the local polynomial used to construct the point estimator from 1 to 2 and 3, respectively. Both remain significant and have similar magnitudes to my main estimate. Finally, Columns (5) and (6) shift the cutoff to 73 and 77 as placebo tests; as expected, neither yields a statistically significant discontinuity.

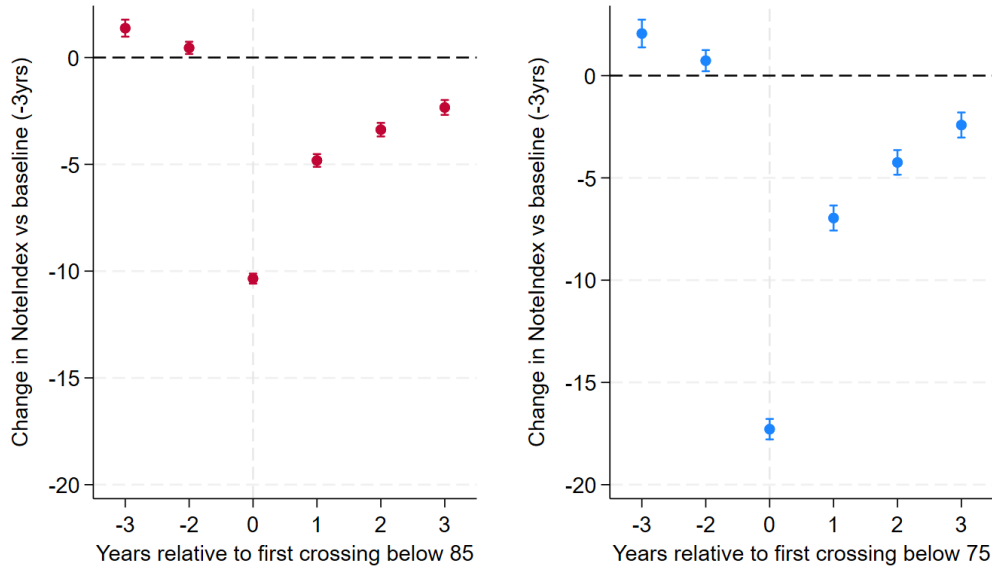
6.2 Event study

Here, I assess what happens to firms as they cross the policy thresholds using an event-time specification with firm and year fixed effects, and controls for headroom and its square root ($\text{headroom}_t = 100 - \text{Index}_{t-1}$), to capture the mechanical ceiling effect and its diminishing returns. Event-time indicators are defined relative to the first crossing year, with $t = -1$ as the omitted baseline year. Table 9 in Appendix B reports results for the 85 and 75 thresholds individually in columns (1) and (2); column (3) includes both sets of indicators to account for firms crossing both thresholds contribute. The combined specification slightly reduces the magnitude relative to the single-threshold columns but preserves the pattern, with all coefficients remaining significant. The results of column (3) are plotted in Figure 1.

Crossing the thresholds at 85 and 75 coincides with a sharp one-year drop in the index: -10.35 points at 85 and -17.29 at 75, relative to the omitted baseline year $t = -1$. Pre-period coefficients are modestly negative, showing there is no large pre-trend of the same order as the $t = 0$ crossing. After the dip, there is a partial rebound; by $t = 3$, the gap has mostly narrowed but not fully to baseline. The 75 threshold shows a larger contemporaneous fall and a steeper recovery than 85. The headroom controls are both highly significant: the linear indicator has a coefficient of 1.013, indicating that for every point below 100, a firm’s growth is increased by 1 point, and the square root of headroom has a coefficient of 0.548, so an additional point of headroom boosts growth more when a firm is near the ceiling (small headroom) and tapers as headroom increases.

Crossings at 75 exhibit a larger fall than at 85, but a noticeably faster recovery. Given this persists when conditioning on headroom, it is unlikely to be driven by mechanical mean reversion. My interpretation is that sub-75 penalties raises the marginal return to improvement,

Figure 1: Event study



Notes: Coefficients are from an event-time specification. $t = -1$ is the baseline year. Dependent variable: change in annual index growth relative to baseline. Vertical bars indicate 95% confidence intervals based on firm-level clustered standard errors. The panel is unbalanced, so firms need not appear in all event years.

inducing quicker post-shock improvements relative to 85.¹⁸

6.2.1 Robustness

Crossing a threshold is not randomly assigned. Firms typically cross thresholds because of a contemporaneous shock, a reorganisation, demand shock, etc., that also moves the index. The event-time coefficients therefore blend the shock that caused the crossing, and additional responses to being below the threshold (reputational, regulatory, etc.). The sharp dips at $t = 0$ are informative about timing, but should be read as descriptive dynamics rather than clean causal effects.

A second issue is the role of headroom. By construction, headroom captures the mechanical ceiling effect, however, starting at $t \geq 1$, headroom is partly an outcome of the event (the $t = 0$ drop increases next period's headroom). In post-event periods, including headroom can act as a post-treatment mediator, soaking up some of the true rebound that occurs precisely because the firm is now far from the ceiling. Because of this, I report three specifications in Table 10 in Appendix B: (1) linear headroom only; (2) linear headroom plus $\sqrt{\text{headroom}}$ to allow for diminishing marginal effects (the main specification); and (3) no headroom controls. The qualitative pattern—a discrete drop at $t = 0$ and partial recovery—survives across all three.

¹⁸This is suggestive, not causal: selection into crossing and shock composition may differ across thresholds.

Thirdly, the large negative $t = 0$ coefficients suggest crossings follow large, unanticipated shocks, rather than a gradual drift, indicating threshold effectiveness. Mostly flat pre-trends reassure that there is no steady slide, but anticipatory behaviour could create a ‘compressed spring’-type mechanism, compressing pre-event declines, but steepening the $t = 0$ cliff.

I allow for dynamics in index growth by adding lagged dependent variables to the event study. These are reported in Table 11 in Appendix B. Year-to-year changes in the index are persistent, so leaving serial correlation in the error term can overstate precision and load persistence onto the event coefficients. Column (1) adds one lag, column (2) adds a second, which are positive and highly significant, indicating persistence. However, neither specifications change the event-time pattern which remains large and significant (slightly attenuated, as expected, because some multi-year propagation is absorbed by the lags).

Finally, my baseline estimates use an unbalanced panel, potentially allowing compositional changes in the set of firms across event years to influence the result. Thus, I re-estimate the model restricting the sample of firms to those observed every year from $t = -2$ to $t = +2$. Results are reported in Table 4. Column (1) reports estimates for the 85 threshold, (2) for the 75 threshold. Magnitudes are similar to baseline, indicating compositional changes are unlikely to drive the main results, though the post-event rebound is faster than base models.

Table 4: Event study estimates: -2 to $+2$ window

Year relative to	(1) 85 threshold	(2) 75 threshold
-2	-0.327 (0.262)	-0.124 (0.492)
0	-12.35*** (0.230)	-18.03*** (0.423)
1	-4.809*** (0.401)	-4.512*** (0.767)
2	-2.955*** (0.314)	0.036 (0.555)
Headroom	1.007*** (0.0746)	0.763*** (0.107)
$\sqrt{\text{Headroom}}$	0.768* (0.414)	2.820*** (0.780)
Constant	-11.49*** (0.618)	-22.79*** (1.498)
Observations	5,201	2,904

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Outcome variable is annual index growth. Event-time indicators are defined relative to the first crossing year, with $t = -1$ omitted.

Table 5: One-year change in components at threshold crossing ($t = 0$)

Component (Δ index points)	Below 85			Below 75		
	Obs	Mean	Std. dev.	Obs	Mean	Std. dev.
Pay gap (wage equality)	5,056	-2.240	6.077	2,541	-5.223	9.582
Individual raises	1,348	-5.423	6.782	647	-7.929	8.595
Promotions	1,375	-1.920	4.390	640	-2.141	5.161
Raises & promotions (small firms)	3,475	-5.289	7.563	1,802	-7.886	9.960
Maternity-return raises	2,439	-4.305	6.854	1,206	-4.154	7.748
Top-10 earners gender balance	5,056	-0.503	2.837	2,541	-0.716	2.799
Total equality index	5,056	-13.036	8.006	2,541	-18.887	10.514

Notes: Δ is current minus previous year. Sample restricted to firms in the first year they cross the respective threshold. Counts differ by row because some subindices are non-calculable in a given year (per official rules).

6.2.2 Pattern of shocks

I decompose the movements across the thresholds to show which dimensions (pay gap, raises, promotions, maternity, top 10 gender balance)¹⁹ account for the $t = 0$ drop. Table 5 reports summary statistics of subindex growth at $t = 0$ for crossings of 85 and 75, respectively.

Several patterns emerge. First, raise-related components account for the largest share of the contemporaneous drop. The post-maternity raise is the second largest contributor; $\approx 29\%$ of firms experience falling below either threshold saw the majority of their drop being from their maternity score. Thirdly, the pay gap subindex is behind a larger proportion of 75 crossings versus 85 ones. The top-10 gender balance moves little on average. Comparing thresholds, the 75 crossings show uniformly larger adverse movements across components than the 85 crossings, consistent with the bigger drop at 75.

There are two caveats here. First, subindex growth is not always calculable, so the composition sample differs across components. I have provided alternative statistics in Table 12, Appendix B, where I drop years without a full set of subindex growth observations, and the magnitudes remain similar. Secondly, again, the decomposition is descriptive rather than causal: the same underlying shock that pushes a firm below a threshold can simultaneously worsen multiple subindices, so parts move together by construction. However, this does show that shocks at $t = 0$ are not driven by a single metric, but by coordinated drops.

6.3 Employee-level triple difference

Table 6 shows the estimates from the triple difference (DDD) specification in equation 3. The post period is from 2020 to 2022 (pre: 2015-2019). Standard errors are clustered at the cohort level. Column (1) presents a baseline DDD with neither controls nor fixed effects. Column (2) adds controls (age, private sector share, socio-professional category shares, industry shares) and year fixed effects. Column (3) introduces cohort fixed effects, so the Female indicator is absorbed and not estimated. Column (4) includes both cohort and year fixed effects with the

¹⁹See either Section 3 or Appendix A for a complete breakdown of subindex scoring.

Table 6: Triple difference-in-difference specification

	(1)	(2)	(3)	(4)
Female	0.0760*** (0.0290)	0.0224** (0.00918)	0 (.)	0 (.)
Post-2020	-0.000461 (0.0246)	0.00638 (0.00625)	0.156*** (0.00897)	0 (.)
Treated	0.802*** (0.0578)	0.0785*** (0.0140)	0.120*** (0.0183)	0.0330*** (0.00950)
Treated \times Post-2020	0.0664 (0.0612)	-0.0206 (0.0154)	-0.188*** (0.0219)	-0.0420*** (0.0123)
Female \times Treated	-0.566*** (0.0698)	-0.0515*** (0.0179)	-0.0831*** (0.0237)	-0.0241* (0.0130)
Female \times Post-2020	0.0849*** (0.0321)	0.0117 (0.00926)	-0.0253* (0.0144)	-0.00877 (0.00770)
Female \times Treated \times Post-2020	-0.151** (0.0760)	0.0138 (0.0215)	0.0716** (0.0321)	0.0373** (0.0177)
Observations	26,398	26,398	26,222	26,222
Controls		✓		✓
Cohort FE			✓	✓

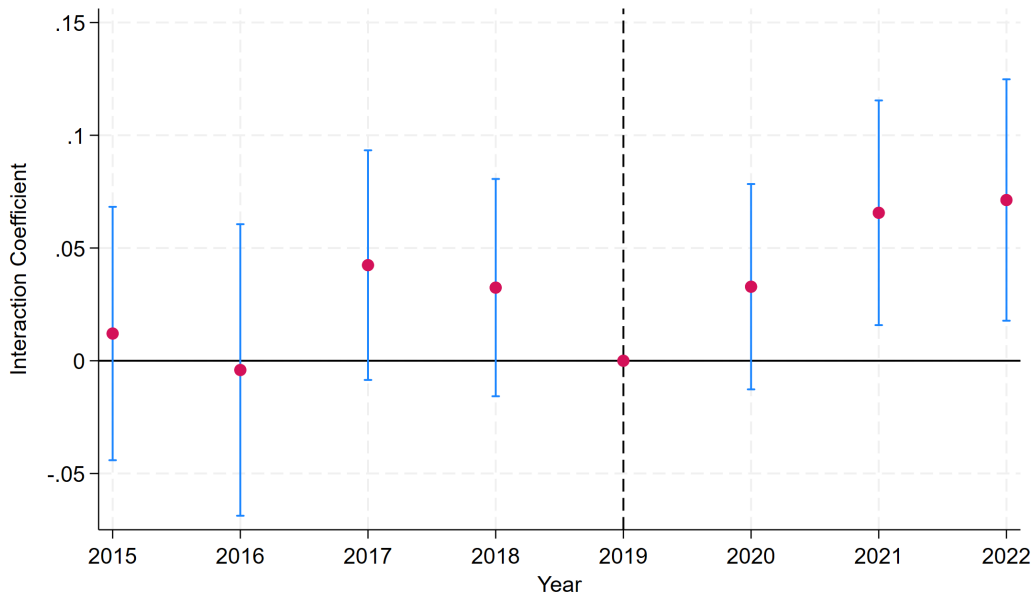
Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable is log hourly wages. Column (1) is the base triple-difference specification; Column (2) adds controls (age, proportion of cohort in the private sector, job-field shares, industry shares); Column (3) uses cohort and year fixed effects with no controls; Column (4) adds controls to the fixed-effects specification. Standard errors in parentheses, clustered by cohort.

controls; as a result, both Female and Post are collinear with the fixed effects and are omitted from estimation.

The final row reports the triple interaction, which is the policy’s effect on the gender wage gap—i.e., the change in female wages relative to male wages in treated firms after 2020. Column (1), which omits both controls and fixed effects, yields a large negative estimate, and a positive coefficient on Female. This is not evidence that women earn more on average; it reflects composition. At the cohort-level, Female compares average wages of female cohorts to male cohorts, and, without fixed effects or controls, any systematic differences in where those cohorts are located in the wage distribution (e.g. from occupational segregation) load onto the Female dummy, a Simpson’s paradox-type aggregation bias.²⁰ Hence, these specifications capture between-cohort differences, not within-cohort changes, so their signs can be

²⁰As a sanity check, individual-level regressions (Table 8, Section 6.3.1) yield a negative coefficient on Female, confirming that the positive cohort-level estimate in (1) and (2) reflects aggregation/composition rather than a reversed gap.

Figure 2: Treatment effect over time



Notes: Dependent variable is real log hourly earnings. 95% confidence intervals shown based on standard errors clustered by firm. Control variables are age, proportion of cohort within the private sector, a given socio-professional category, and a given industry. Year and cohort fixed effects are included. Figure based on estimates from Table 14 in Appendix B.

misleading.

Column (3) is an estimation with cohort fixed effects. The triple interaction becomes statistically significant, indicating a narrowing of the gender gap within cohorts over time. Column (4), which includes both cohort and year fixed effects alongside controls, is my preferred specification. This results in an estimate of the policy effect of 0.037, significant at the 5% level. Given the dependent variable is the log of hourly wages, this can be interpreted as roughly a 3.7 percentage point increase in female hourly wages relative to male hourly wages. The negative coefficient on the Treated \times Post interaction suggests that the effect is driven by a reduction in the pay of male workers, consistent with previous literature (Blundell, 2021; Bennedsen et al., 2022).

Figure 2 plots the year-by-year triple interaction coefficients from specification 4, with error bars representing 95% confidence intervals. Coefficients are normalised to zero in 2019, the last pre-policy year. Pre-2019, all estimates are statistically insignificant, supporting parallel trends. Post-policy, the 2020 coefficient is also insignificant, while 2021 and 2022 are positive and significant, suggesting that the policy took some time to gain effectiveness. This contrasts Bennedsen et al. (2022) and Blundell (2021), in equivalent graphs for the British and Danish policies, where coefficients are significant immediately, suggesting anticipatory wage adjustment; by contrast, my estimates are consistent with firms responding after the first published reports in 2020. The full set of estimates are in Table 14 in Appendix B.

In Table 7, I estimate alternate specifications: columns (1) and (2) estimate log annual wages

Table 7: Alternative outcomes and samples: triple-difference estimates

	(1) Log annual wage	(2) Log hours	(3) Log hourly (FT+PT)	(4) Log annual (FT+PT)
Treated	0.0333*** (0.00981)	-0.103** (0.0442)	0.0477*** (0.0142)	0.0531*** (0.0147)
Treated \times Post-2020	-0.0501*** (0.0126)	-0.0717 (0.0522)	-0.0929*** (0.0158)	-0.104*** (0.0167)
Female \times Treated	-0.0294** (0.0137)	0.105 (0.0662)	-0.0705*** (0.0184)	-0.0804*** (0.0191)
Female \times Post-2020	-0.0154* (0.00804)	0.0605* (0.0355)	-0.0524*** (0.0103)	-0.0621*** (0.0108)
Female \times Treated \times Post	0.0540*** (0.0185)	0.00295 (0.0817)	0.133*** (0.0228)	0.154*** (0.0240)
Observations	26,222	26,222	27,878	27,878
Average cohort size	163.30	163.30	194.02	194.02

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Column (1) uses annual log wages as the outcome; (2) uses log hours; (3) includes both part-time and full-time workers with log hourly wage as the outcome; (4) repeats this using annual log wage. All specifications are estimated in cohorts with the triple-difference design. Standard errors in parentheses, clustered by cohort.

and log hours, respectively, on the full-time workforce, while (3) and (4) also include the part-time workforce in estimations of log hourly wages and log annual wages. For annual specifications, I include the log of hours as an additional control. In my original full-time only setup, part-time jobs are excluded before cohort construction. To include them, I reconstruct the cohorts using the same procedure, which increases the average cohort size and reduces the number of small cohorts (< 30 individuals) being dropped.²¹

The results show a positive and significant triple-interaction effect in columns (1), (3), and (4). The coefficients for the full workforce in (3) and (4) are larger than for the full-time-only sample, consistent with women in part-time jobs having lower baseline wages and thus magnifying the observed treatment effect. This pattern implies that women in full-time employment are positively selected relative to the overall female workforce. This occurs without an equally large decrease in the Treated \times Post term, suggesting that here, women's wages increased alongside a fall in men's wages. Column (2) shows now detectable effect on hours worked, implying that effects are driven by wages, not changes in the labour supply.

²¹It is implausible that people do not transition between part- and full-time jobs, so this cannot be used in the cohort definition, meaning cohorts include both part- and full-time workers.

6.3.1 Robustness

Table 8 reports various robustness tests. Column (1) removes the square root of cohort-size weights used in all other cohort specifications, producing a slightly larger and still significant effect. The second column drops all cohorts that are smaller than 100 individuals, dramatically reducing the sample size; the coefficient remains similar in magnitude but loses significance. Column (3) fixes all control variables at their 2019 values in the post-treatment period (2020-2022) to address potential endogeneity if the policy affected control variables; the triple interaction remains highly significant. The fourth and fifth columns present individual-level OLS regressions.²² Column (4), estimated without controls, shows no significant triple-interaction but it does find a significant negative gender pay gap of 0.128 (roughly 13.7%). Column (5), which adds controls for age, a private sector dummy, socio-professional category, industry, and year fixed effects, yields a positive and highly significant triple-interaction, though with a far smaller magnitude than in cohort-level specifications, though it is roughly one third of the estimated pay gap. In both, unlike the cohort-level results, the Treated \times Post coefficient is also positive and significant.

Table 13 in Appendix B reports results from alternative pseudo-panel cohort definitions. Columns (1) and (2) group birth years in 3- and 4-year bins, respectively, instead of 2-year bins, as well as sex and department. (3) replaces department with regions and uses exact year bins. Specification (4) defines cohorts by region, six industry groups, and 2-year birth bins.²³ While point estimates remain similar in sign and magnitude to baseline specifications, they are not statistically significant.

This loss of precision is mechanical: coarser cohorts reduce within-cohort homogeneity, attenuating estimated effects, while finer or cross-classified cohorts produce many smaller cells, which are then dropped for being too small, reducing the effective sample size and representativeness. Variables such as age and sex have smooth distributions, but region/department and industry are highly unbalanced. Combining multiple heterogeneous variables yields many small cohorts and a long tail of large cohorts. For example, any small region paired with a small- or medium-sized industry will be dropped. Figure 4 illustrates this in specification (4), where most cohorts lie just below the inclusion threshold and a long right tail contains a handful of extremely large cohorts.

Overall, these robustness checks suggest that the baseline results are not driven by a particular cohort definition. The direction of the coefficients are stable across specifications, and the lack of significance is consistent with the expected trade-off between cohort granularity and estimation precision in pseudo-panel analysis.

7 Conclusion

As gender pay gap legislation is being implemented in an increasing number of countries, it is useful to assess how effective existing policies are and whether threshold mandates, like those in France, can alter firm behaviour. I find that the French policy reduced the gender pay gap

²²Here, ‘Treated’ becomes a dummy, rather than a cohort-level proportion.

²³This requires a strong assumption that individuals do not change industries. Remember, this isn’t ‘type of job’, but rather what the firm they work for does.

Table 8: Triple difference-in-differences specification robustness checks

	(1) No Weight	(2) Large Cohorts	(3) Fixed Controls	(4) Indiv. Level	(5) Indiv. Level
Female	0 (.)	0 (.)	0 (.)	-0.128*** (0.000721)	-0.0772*** (0.000596)
Post-2020	0 (.)	0 (.)	0 (.)	0.0231*** (0.000797)	0.00330*** (0.000841)
Treated	0.0336*** (0.00993)	0.0330** (0.0137)	0.0312** (0.0121)	0.0387*** (0.000785)	0.0184*** (0.000609)
Treated \times Post-2020	-0.0442*** (0.0132)	-0.0546*** (0.0182)	-0.124*** (0.0179)	0.00395*** (0.00126)	0.00359*** (0.000973)
Female \times Treated	-0.0202 (0.0139)	-0.0432** (0.0186)	-0.0254 (0.0163)	-0.0302*** (0.00111)	-0.0102*** (0.000858)
Female \times Post-2020	-0.0142* (0.00808)	-0.00632 (0.0113)	-0.0216** (0.0102)	0.0225*** (0.00116)	0.0163*** (0.000895)
Female \times Treated \times Post	0.0498*** (0.0188)	0.0327 (0.0258)	0.0674*** (0.0239)	-0.000984 (0.00178)	0.00811*** (0.00137)
Observations	26,222	14,939	25,720	4,367,476	4,364,634
Controls	✓	✓	✓		✓
Cohort FE	✓	✓	✓		

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: log hourly wage. Columns (1)–(3) are cohort-level specifications with cohort-clustered standard errors: Column (1) removes cohort-size weighting, Column (2) drops cohorts below 100 individuals; Column (3) uses variables frozen to 2019 levels during the post-2020 period. Columns (4)–(5) are individual-level OLS, where “Treated” is a dummy, not a cohort-level proportion; Column (5) includes controls of age, a private sector dummy, socio-professional factor variables, industry factor variables, and year fixed effects.

by 3.7 percentage points amongst affected firms, with the adjustment largely coming from lower male wages. However, a limitation with my approach was that I considered only firms around the 50-employee threshold: effects could differ among larger firms.

I find that firms just below the 75 penalty threshold grow 1.6 index points per year faster than those just above, consistent with incentives to avoid fines. Event-time estimates show that crossings are typically driven by large one-period negative shocks, not gradual drift. Those falling below 75 experience a bigger drop than those crossing 85, yet recover faster, suggesting fine-avoidance as an effective motivation. While France is reforming this legislation, abandoning thresholds would risk weakening incentives to improve gender pay gaps.

Future work would benefit from the full specification of the *Base tous salariés* that forms a true panel, gives granular employer size to enable an RDD, and, especially, provides specific firm identification (SIREN). This would allow firm-level controls in a triple-differences strategy and a more granular examination of threshold behaviour. One could test potential index exploitability—for example, a firm could give broad but small raises to lift scores—by comparing reported subindices to underlying pay movements.

References

- Baker, M., Halberstam, Y., Kroft, K., Mas, A., and Messacar, D. (2023). Pay transparency and the gender gap. *American Economic Journal: Applied Economics*, 15(2):157–83.
- Bennedsen, M., Simintzi, E., Tsoutsoura, M., and Wolfenzon, D. (2022). Do firms respond to gender pay gap transparency? *The Journal of Finance*, 77:2051–2091.
- Bertrand, M., Kamenica, E., and Pan, J. (2015). Gender identity and relative income within households *. *The Quarterly Journal of Economics*, 130(2):571–614.
- Bissuel, B. and Métais, T. (2025). France seeks new rules to narrow the gender pay gap. *Le Monde (English)*. In-depth consultation with unions and employers launched to transpose EU pay transparency directive.
- Blau, F. D. and Kahn, L. M. (2000). Gender differences in pay. *Journal of Economic Perspectives*, 14(4):75–99.
- 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. (2021). Wage responses to gender pay gap reporting requirements. CEP Discussion Papers dp1750, Centre for Economic Performance, LSE.
- Blundell, J., Duchini, E., Simion, S., and Turrell, A. (2025). Pay transparency and gender equality. *American Economic Journal: Economic Policy*, 17(2):418–45.
- Böheim, R. and Gust, S. (2021). The Austrian pay transparency law and the gender wage gap. Technical report, IZA - Institute of Labor Economics.
- Cahuc, P. (2024). Labour market inequality in France and in the UK. *Oxford Open Economics*, 3(Supplement_1):i933–i939.
- Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.
- Cattaneo, M. D., Jansson, M., and Ma, X. (2018). Manipulation testing based on density discontinuity. *The Stata Journal*, 18(1):234–261.
- Caudill, C., Farruggia, A., Sonnenholzner, L., Stobbe, S., and Wirsching, K. (2023). *The Gender Pay Gap in France*, pages 175–193. Springer Nature Switzerland, Cham.
- Chetrit, J. (2024). Index de l'égalité professionnelle femmes-hommes : un outil décevant et une réforme qui se profile. *Le Monde*. Publié le 07 mars 2024 à 07h00, modifié le 07 mars 2024 à 09h33.
- Clarke, D. and Tapia-Schythe, K. (2020). Implementing the panel event study. IZA Discussion Paper 13524, IZA Institute of Labor Economics, Bonn, Germany.
- Connolly, H. (2024). The gender equality index in France: what is it and how does it work? Politics of Equality at Work blog, University of Manchester. Accessed: 2025-08-24.

- Cullen, Z. and Perez-Truglia, R. (2018). How much does your boss make? The effects of salary comparisons. Working Paper 24841, National Bureau of Economic Research.
- Cullen, Z. B. and Pakzad-Hurson, B. (2023). Equilibrium effects of pay transparency. *Econometrica*, 91(3):765–802. Originally circulated as Working Paper, June 2021.
- Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, 30(1-2):109–126.
- European Commission (2022). 2022 factsheet on the gender pay gap. https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/equal-pay/gender-pay-gap-situation-eu_en. PDF document, accessed 2025-09-08.
- Galanakis, Y. and Gosling, A. (2024). Mind the (Gender Pay) Gap - The role of Board Gender Composition. Working Papers 045, The Productivity Institute.
- Garicano, L., Lelarge, C., and Van Reenen, J. (2016). Firm size distortions and the productivity distribution: Evidence from France. *American Economic Review*, 106(11):3439–79.
- Gobillon, L., Meurs, D., and Roux, S. (2015). Estimating gender differences in access to jobs. *Journal of Labor Economics*, 33(2):317–363.
- Groshen, E. L. (1991). The structure of the female/male wage differential: Is it who you are, what you do, or where you work? *The Journal of Human Resources*, 26(3):457–472.
- Guillerm, M. (2015). Les méthodes de pseudo-panel. Documents de Travail de l’Insee - INSEE Working Papers m2015-02, Institut National de la Statistique et des Etudes Economiques.
- Guillerm, M. (2017). Pseudo-panel methods and an example of application to household wealth data. *Economie et Statistique / Economics and Statistics*, None(491-492):109–130.
- Gulyas, A., Seitz, S., and Sinha, S. (2023). Does pay transparency affect the gender wage gap? Evidence from Austria. *American Economic Journal: Economic Policy*, 15(2):236–55.
- INSEE (2025). Gender pay gap in 2023. Insee Focus no. 349. Women’s average wage income was 22.2
- International Labour Organization (2009). Global employment trends for women. International Labour Office, Geneva. Accessed: 2025-08-28.
- International Labour Organization (2025). Employment-to-population ratio by sex and age (%) – annual, France. ILOSTAT Data Explorer. LFS – Employment Survey data. Accessed: 2025-08-28.
- Jones, M. K., Kaya, E., and Papps, K. L. (2023). The ongoing impact of gender pay gap transparency legislation. IZA Discussion Paper 15817, Institute of Labor Economics (IZA).
- Mas, A. (2014). Does transparency lead to pay compression? Working Paper 20558, National Bureau of Economic Research.

- Meng, Y., Brennan, A., Purshouse, R., Hill-McManus, D., Angus, C., Holmes, J., and Meier, P. S. (2014). Estimation of own and cross price elasticities of alcohol demand in the UK—a pseudo-panel approach using the Living Costs and Food Survey 2001–2009. *Journal of Health Economics*, 34:96–103.
- Ministry for Europe and Foreign Affairs (2021). Gender equality at the Ministry for Europe and Foreign Affairs. French Ministry for Europe and Foreign Affairs. Accessed: 2025-08-28.
- Nergaard, E. (2002). Gender pay equity in europe. Retrieved on 03 Sep 2025.
- Office for National Statistics (2022). Percentage of employees by enterprise size. Office for National Statistics, Reference number 14128. Accessed: 2025-08-24.
- Schäfer, A. and Gottschall, K. (2015). From wage regulation to wage gap: how wage-setting institutions and structures shape the gender wage gap across three industries in 24 european countries and germany. *Cambridge Journal of Economics*, 39(2):467–496.
- Verbeek, M. (2008). *Pseudo-Panels and Repeated Cross-Sections*, pages 369–383. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Verbeek, M. and Nijman, T. (1992). Testing for selectivity bias in panel data models. *International Economic Review*, 33(3):681–703.
- Verbeek, M. and Nijman, T. (1993). Minimum MSE estimation of a regression model with fixed effects from a series of cross-sections. *Journal of Econometrics*, 59(1-2):125–136.
- World Economic Forum (2024). Global gender gap report 2024. Technical report, World Economic Forum, Geneva, Switzerland. 18th edition.
- World Economic Forum (2025). Global gender gap report 2025. Technical report, World Economic Forum, Geneva, Switzerland.