
Impacts of Equal Pay Acts and Pay Transparency on gender
discrimination on labour market:

An adventurous journey in DiD estimation.

Camille Souffron

Ecole Normale Supérieure

camille.souffron@ens.psl.eu

Contents

1	Introduction	2
2	Theoretical framework(S) and literature review	2
3	Equal pay policies and Pay Transparency	3
4	Data and descriptive statistics	5
4.1	Which output variables should be considered?	5
5	Theoretical approach: On the difficulties of DiD estimation for the isolated net effect of pay transparency	6
5.1	TWFE/DiD with two groups and two treatments?	8
5.2	Canonical TWFE/DiD with three groups and two treatments	9
5.3	Staggered DiD with several treatments	10
6	DiD between Vermont and Illinois on Pay Transparency	13
7	Finally, back to a trivial DiD: estimating the cumulative effect of the 2003 Act	15
7.1	Model	16
8	Results	19
8.1	Fixed Effects OLS Results	19
8.2	SMC results	22
8.3	Placebo results	23
9	Discussion	23
9.1	Non-random selection	25
9.2	On the endogeneity and measurement issues of human capital	26
9.3	On the dimensions considered	27
9.4	On wage distribution	27
10	Conclusion	28
	References	29

1 Introduction

NB: the code in R is available on request, as are the DiD and OLS tables for robustness checks (e.g. with log-difference of mean wages. This work was carried out independently, without any discussion with the course leader - due, obviously, to the author's sole fault and health contingencies. The work surely does not fall within the scope of what is expected, but at the very least it attempts to demonstrate as much critical thinking as possible about the possibilities of autonomously answering the research question, by explaining all the stages of an independent reasoning.

On average, women are paid about 20 per cent less than men, globally (ILO, 2019). In the United States in 2016, on a GWG of 18 and 20 percent at the median of the wage distribution and between 21 and 24 percent at the mean (Foster et al., 2020). In the 75 years since the end of World War II, the experience of women in the American labor market has changed. Seven out of 10 women are now employed, making up almost half of the workforce. They spend more time in their professional life and correspondingly less time in purely nursing activities. Motherhood is no longer incompatible with paid work. Women have outperformed men academically for decades. However, the labor market experiences of men and women remain different. Women continue to be employed primarily in the service and public sectors, as well as in the nursing profession. They make up more than four-fifths of part-time workers but remain underrepresented in top and management positions. Despite some convergence in wages between men and women, there is still a considerable gap. Several explanatory theories for this gap exist in the economic literature.

2 Theoretical framework(S) and literature review

Traditional neo-classical theory has explained GWG in terms of gender differences in the rational accumulation of human capital since Becker (1962) and Mincer (1974 for his human capital income function). This human capital is the level of education coupled with the level of work experience - with women investing less in education (which is false nowadays, as women in the USA, as on average in developed countries, outperform men at tertiary level but now also at secondary level (www.census.gov)) and working less by choosing to bring up their children or to be housewives (this may seem caricatural, but it is *stricto sensu* Becker's theory of the division of labour according to comparative advantages in the household (1985). Human capital is a proxy for the differences in marginal productivity, which is supposed to be the factor that determines real wages.

However, human capital only explains around half of gross GWG. In the United States in 2016, out of a GWG of 18 and 20 percent at the median of the wage distribution and between 21 and 24 percent at the mean (Foster et al., 2020), studies that control for human capital (including work experience), occupation, and industry, which is a substantial number of characteristics, are unable to explain between 38 and 71 percent of the gender wage gap (CONSAD 2009; Chamberlain 2016; Blau and Kahn 2017). Although the gender wage gap has narrowed over the years, a growing share of the remaining gap is unexplained (*Ibid.*).

Moreover, human capital explanations of the GWG have declined in importance over time - although they still contribute to the gender wage gap - Goldin (2014); Blau and Kahn (2017). The net or residual unexplained GWG, i.e. controlled for human capital, is therefore caused by other differences whose role is growing, and which should be reduced by public policies to allow convergence modulo human capital¹. Thus, a number of theories aim to explain net or residual GWG: statistical (Arrow, 1973) or taste-based (Becker, 1957) discrimination by the employer, for example due to the perceived risk of discontinuity (Lazear and Rosen, 1990), gendered preferences in job-seeking, division of labour in the household, monopsony and segmentation in the labour market with female sectors paying less (Webber, 2016; Barth et al., 2016), negotiation of lower wages in exchange for more flexible working hours for women to manage *care responsibilities* (Goldin, 2014), monopsony power and differential in wage bargaining capacity (Artz et al., 2018, find women ask but they don't get), women would be less competitive (lab experiments say yes (Gneezy et al. 2003; Niederle and Vesterlund, 2007) but not in real data (Manning and Saidi, 2010; Bryan and Bryson, 2016)), psychological differences (e.g. risk averse "account for small to moderate amount of gap" (Blau and Kahn, 2017), fatherhood premium and motherhood penalty (Costa Dias et al., 2020; Yu and Hara, 2021) et caetera.

3 Equal pay policies and Pay Transparency

Faced with this residual non-trivial GWG, policies based on the law and mandatory nature, particularly of equal pay, have limits and legistic enunciation does not necessarily imply performativity, and even if the GWG has continued to decline since 1980, the convergence has slowed, as in other developed countries (Kunze, 2017): despite equal pay legislations of 1963², 1971 and 1972, the GWG in 2019 was still 18% in the US (Polachek, 2019), 20.8% in the UK, 16% in France (Eurostat, 2019), and 12.7% in 2021 in the EU27 (Eurostat, 2021) - those values are gross GWG.

Thus, if the gender wage gap continues to shrink at the rate it has between the passage of the Equal Pay Act in 1963 and 2021, median full-time working women will not achieve pay parity with men before 2056 - it will take even longer for working women of colour (**CAP, 2021**). This is despite the federal law of 1963 and its recent widespread application in State laws (except in Missouri). Equal pay law procedures also require a comparison of pay for a man and a woman in similar jobs - there is not always anything to compare them with. What's more, despite the illegal nature of this discrimination, it is difficult for a woman to take legal action because of the psychological, emotional, financial and signalling costs involved (for example, in the UK, litigation and collective bargaining are necessary complements, Deakin et al., 2015). Finally, discrimination in recruitment segments sectors beforehand, upstream of the context of application of equal pay laws (problem if unequal access to jobs in first place).

So all this calls into question the possibility of doing away with the residual GWG simply by comminatory legislation. An alternative strategy is to require transparency in the way employers pay men and women. Some scholars argue that pay secrecy can contribute to the gender pay

¹We will see in the **Discussion** that human capital and "gendered rational preferences" are themselves endogenous to gender inequalities and restrictive social norms.

²Section 206(d)(1) of the Federal Labor Act: <https://www.eeoc.gov/statutes/equal-pay-act-1963>

gap (Eisenberg, 2011) because it can help companies “avoid perceptions of unfairness when pay inequities do exist and can minimize claims of discrimination” (Colella et al. 2007). It can affect employers’ practices because those seen to be presiding over large GWGs are liable to suffer reputational damage in the eyes of consumers, investors, and potential employees. And it can affect workers and job searchers by providing them with relevant information and evidence that they require to negotiate pay rates and provide them with the means to challenge potential pay discrimination. There is evidence from other countries that pay transparency laws have helped reduce the GWG in Denmark (Bennedsen et al., 2018) and Switzerland (Vaccaro, 2017). Many European countries already have pay transparency legislations in place, with an EU-wide Pay Transparency Directive pending adoption, and non-European countries, such as Canada, Chile, Iceland, have passed such laws. It therefore seems appropriate to study the impact of such a pay transparency policy, in isolation from equal pay legislation, on the GWG, but also to study whether there are any negative effects on demand for women’s work, in this case in the United States.

Since 1997, various US Congresses have attempted to pass the Paycheck Fairness Act, which would impose such a pay transparency policy at federal level. Still rejected by the US Senate in June 2021, it would have allowed the Equal Employment Opportunity Commission to regulate the data collection from employers. Furthermore, the Act would have banned employers from asking job candidates their previous salary, guaranteed employees the right to discuss their salaries without fear of retaliation, and extended the remedies available to victims of discrimination on race or ethnicity grounds to gender discrimination cases, among other provisions³.

Nevertheless, several states have enacted pay transparency laws recently or are considering pay transparency laws (e.g. California, Colorado, Washington in 2021, Maryland in 2020) so that 22 states already have such laws. Illinois was the first state to pass a local pay transparency law, in 2003, which was implemented in 2004 as part of the Equal Pay Act, the local version of the 1963 federal law. Illinois is also the second state to have introduced a local amendment to the Equal Pay Act, after Indiana in 1965. The “pay transparency” section states that *“It is unlawful for any employer to discharge or in any other manner discriminate against any individual for inquiring about, disclosing, comparing, or otherwise discussing the employee’s wages or the wages of any other employee”* (820 Ill. Comp. Stat. 112/10(b)). It applies to all employees and employers, including the state. This situation therefore provides us with an ideal empirical framework for comparing the effects of the policies under consideration with the other American states.

Note that there are many possible pay transparency measures (ILO, 2022): from allowing employees to request and access information on pay levels in their enterprise to requiring employers to disclose individual pay information to employees, as well as prohibiting employers from requesting an employee’s or prospective employee’s salary history, creating an independent body to provide employers with equal pay certification if they meet certain requirements around gender-neutral pay, and even obliging enterprises with a certain threshold level of employees (for example, 50) to publish information on gender and pay within their organization. The pay transparency introduced by Illinois is therefore minimal: it simply implies a prohibition on dis-

³The text of the bill is available at: <https://www.congress.gov/bill/117th-congress/house-bill/7/text>

crimination/sanctioning the research and disclosure of salaries by employees, and can therefore fall within the scope of theories of discrimination in hiring but also in promotion, and in the theory of salary negotiation. It is this, together with the local Equal Pay Act, that we will be assessing.

4 Data and descriptive statistics

The data used is pooled individual data (not panel data) from the very exhaustive US Census Population Survey (CPS) from 2000 to 2006, i.e. over 7 years (this will be important), sectorised for the 50 States and the District of Columbia using their ZIP code, but not by city. Each individual has a household ID and a multinomial variable *relate* indicating his or her place in the household - all household members are questioned. This makes it possible to reconstruct different statistical entities, from the household (allowing "within" regression) to the State by collapsing the data (which will be our first approach as an example, before showing its extreme limitation in terms of statistical power and type II error).

The sample comprises 344,287 observations - 168,376 women and 175,911 men - over 7 years, described using 28 variables in addition to the year and the household ID (sex⁴, age, nominal and real hourly wages, professional status, weekly working hours, etc.), variables which will provide controls (marital status, number of individuals in the household, duration and level of education, ethnic and trade union affiliation, etc.).

We restrict the population studied to people aged between 25 and 50 in order to focus on the main period of activity, and we also restrict ourselves to people employed in the private sector (neither public nor self-employed, as the public sector is more likely to respect equal pay.), leaving us with a sample of 237,835 observations (110,679 women and 127,156 men). In collapsed data, this provides 51*7 state-year cells (51 because of Washington DC), offering a panel. Finally, we exclude Indiana because, according to our own research based on the instructions, it was the first state to pass an Equal Pay law before Illinois (and did so as early as 1965, preventing a robust or useful comparison).

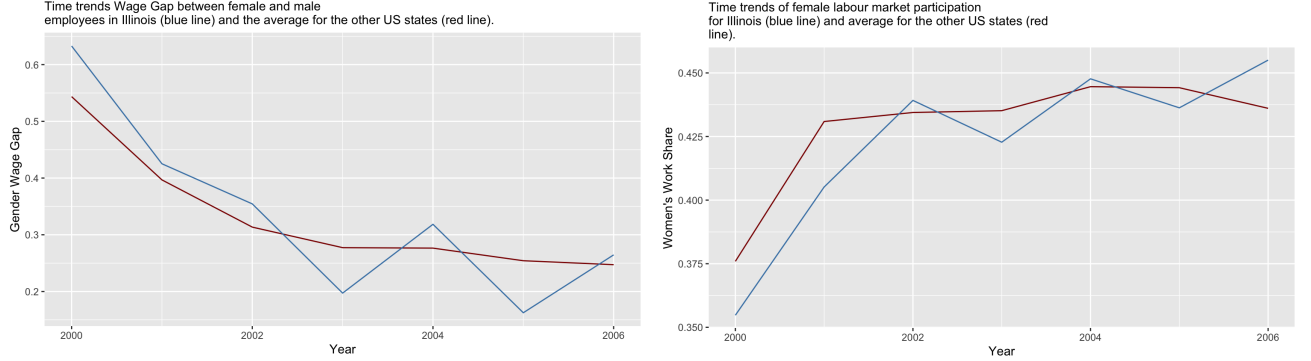
4.1 Which output variables should be considered?

Our object of interest here is mainly GWG, particularly net GWG (controlled for human capital). However, in a holistic analysis of gender inequalities in the labour market, and in order to study inclusive policies, we will also look at the women's work share. Traditional economic theory assumes an increase in unemployment in the event of a rise in wages, particularly affecting workers on the lowest incomes, as well as a recomposition of the combination of factors of production in the event of an increase in the relative price of one (Hamermesh, 1993; Buchanan, 1996; Neumark et al., 2004).⁵ We could therefore imagine a negative impact on female employment if policies to reduce GWG were strengthened, with employers employing fewer women.

⁴the data present will not allow us in this case to distinguish between biological sex and social gender

⁵Although this has been refuted by empirical studies (e.g. the literature around Card & Krueger, 1994). It should be remembered that the use of the empirical approach to refute or at least qualify an "iron law of economics" derived from "common sense" led Buchanan to say that some economists, including Card & Krueger, had become "camp-following whores", en français des « prostituées de bordel militaire de campagne ». Le correcteur économètre appréciera.

It should be noted that in the UK, neither the Equal Pay Act of 1971, nor the introduction of a minimum wage in 1999, two policies which considerably reduced the GWG, had any detrimental effect on demand for women labour, even low-waged one (Zabalza & Tzannatos, 1985; Robinson, 2002). Note that in terms of human capital theory, this suggests that women had been paid below their marginal product and systematically undervalued.



We can see a downward trend in GWG in the U.S. states over the period 2000-2006, but with an ever-smaller marginal decline, the main drop taking place between 2000 and 2002 (GWG is calculated here in log-difference c.f. *infra*). Illinois experienced this reduction, with a significant drop between 2002 and 2003, before rising again the following year, then falling again in 2005 and again in 2006, thus experiencing significant variations. As for the women's work share, it rose sharply in 2000 for the US states (and in 2001 for Illinois), before stabilizing and even declining slightly from 2005 onwards on average. Illinois, on the other hand, saw an increase in 2005.

5 Theoretical approach: On the difficulties of DiD estimation for the isolated net effect of pay transparency

Thus, we are interested in "the impact of the pay transparency act on gender equality in the labor market" in the context of Illinois, using US state census data for the period 2000-2006 ($N = 7$). Modulo the questions of external and ecosystemic validities as well as transposability, the identification and estimation of the impact of the pay transparency act in Illinois would make it possible to approximate the more general impact of pay transparency policies on gender equality in the labor market, in comparison, as we have seen *supra*, with imposed equal pay policies which do not make it possible to eliminate the gender pay gap.

As stated, we assume that public policies to reduce gender pay inequalities are part of a broader approach to reducing gender discrimination in the labour market, an approach that is intended to be coherent. Thus, in addition to studying the impact of the policy on the reduction of the wage gap, we also study its impact on the female labour participation rate, in order to investigate the hypothesis of a negative effect of the reduction of the wage gap on this rate. Our two outcomes of interest $y_{s,t}$, collapsed by State and at each period, are :

- gender pay gap, calculated as the difference in log median real (in 2010 dollars) hourly

wages for men and women, corresponding to the ratio of the two wages for men to women. Medians to reduce the impact of extreme values (The estimates that follow in this work have also been tested for GWG mean values as a robustness test, and the results are equivalent). This value is "how many times more a man earns than a woman", and the median female wage would have to be multiplied by 1 minus this value to reach the median male wage. Thus, the estimated impact of the complete policy is $(\beta_7 * 100)\%$ on the wage gap rate:

$$\text{gender wag gap} = \log(\text{men's median wage}) - \log(\text{women's median wage})$$

- As well as the work share of women in the population:

$$\text{women's work share} = \frac{\text{hours worked by women employee per week}}{\text{total worked hours by the population per week}}$$

And here come the problems. Which design to use?

Here we seek to isolate the effect of the pay transparency policy on two outcomes related to the gender gap in the labor market. According to the available Bachelor thesis article (Andersson and Ekman, 2018), Illinois "was one of the first (the second) states to implement its own amendment to the 1963 federal equal pay act", the 2003 Equal Pay Act, which prohibits gender wage discrimination in the private sector. This state law also incorporates a pay transparency policy, the third in American history after Michigan (1982) and California (1984): *"It is unlawful for any employer to discharge or in any other manner discriminate against any individual for inquiring about, disclosing, comparing, or otherwise discussing the employee's wages or the wages of any other employee."* (820 Ill. Comp. Stat. 112/10(b)). It is thus a cumulative policy of equal pay and pay transparency.

It seems appropriate to use the Difference-in-Difference (DiD) method, which is both the most common and the oldest quasi-experimental research design, dating back to Snow's (1855) analysis of a London cholera outbreak. A DD estimate is the difference between the change in outcomes before and after a treatment (difference one) in a treatment versus control group (difference two);. That simple quantity also equals the estimated coefficient on the interaction of a treatment group dummy and a post-treatment period dummy in a regression. Under parallel trends hypothesis, the DD estimate identifies the average treatment effect (ATE).

In all that follows, we use the Two Way Fixed Effects (TWFE) method to estimate the DiD, following Bertrand, Duflo and Mullainathan (2004), to avoid having to collapsed our data into just two periods (pre- and post-treatment), while at the same time avoiding having standard errors and therefore confidence intervals totally overestimated due to the obvious serial correlations between each period⁶. Also, it will be considered for now that the interventions are as good as random, conditional on time and group fixed effects. We will talk about the validity of

⁶Bertrand et al. (2004) use DiD to estimate the impact of "placebo" reforms on women's wages in randomly selected US states, and find 45% of false-positive results to be significant at 5%, instead of the 5% expected at a threshold of 1.96 for t-statistics. This is because of the obvious serial correlations between t and $t+1$ for any t in the state panel data

a DD estimates in the face of possible endogeneity of the interventions themselves.

5.1 TWFE/DiD with two groups and two treatments?

Thus, we can attempt to compare the effects of this dual policy on our outcomes with the policy pursued either in the first state, which has only implemented equal pay, or in a state which will do so at a later date, and isolate the effect of pay transparency (modulo the differences between states and under the hypothesis of common trends).

Nevertheless, a theoretical problem arises. The DiD estimator is unbiased when comparing two groups with different treatments, only if at least one of the two treatments has a constant effect (Fricke, 2015). Indeed, the assumption of parallel trends in the absence of a control group must be rearranged as follows (with P1 the Equal Pay policy and Pol the combined policy):

$$\begin{aligned} & E(Y_{t=2}^0|D = P1) - E(Y_{t=1}^0|D = P1) + E(Y_{t=2}^{Pol}|D = P1) - E(Y_{t=2}^0|D = P1) \\ &= E(Y_{t=2}^0|D = Pol) - E(Y_{t=1}^0|D = Pol) + E(Y_{t=2}^{Pol}|D = Pol) - E(Y_{t=2}^0|D = Pol) \end{aligned} \quad (1)$$

where the first difference for each member corresponds to the common trends in the non-treatment situation, i.e. the effect of time. The second difference imposes the homogeneity of the effect of the combined policy "Pol" in relation to non-treatment for each post-treatment group. Note that in reality there is no "imposition", as there may be heterogeneity of effect, but only if a deviation of the Common Trends under non-treatment perfectly compensates for it - the fact remains, this counterfactual would not be observable, so homogeneity is required for identification. This constraint is all the stronger when we have to impose this hypothesis conditionally on the covariates, as it must resist unobservables - yet homogeneity seems far too restrictive, even conditionally, as many groups are not treated randomly but because of observable and unobservable characteristics (confoundedness problem). Indeed, given the type of policies studied, it seems unlikely to assume a conditional independence assumption (i.e. non-selection observable)⁷. Nevertheless, Fricke (2015) demonstrates that under the assumptions of effects of the same sign for both treatments and that one effect is always strictly greater than the other for both groups, the DiD estimator estimates at least a lower bound in absolute values on the average treatment effect on the treated compared to the unobserved non-treatment state, even if effect homogeneity is violated.

But we have an alternative: the use of a third group, a control group. Indeed, if we were to carry out a DiD only between Illinois and the group of untreated states, we would only be able to identify and estimate the effect of the 2003 cumulative policy, i.e. Equal Pay plus Pay Transparency. This would only be the interaction term of the two policies, and we could not isolate the effect of P2 i.e. Pay Transparency.

⁷e.g. Duflo (2001) controls for pre-treatment levels of schools in one specification, which are arguably an important dimension of effect heterogeneity. However, the level of schools also perfectly determines the regions' treatment status. Hence, common support is violated by construction

5.2 Canonical TWFE/DiD with three groups and two treatments

However, with the other 48 states over this period, we can largely construct an untreated baseline. To study this, let's start with a simple framework, which is the canonical DiD framework (simultaneous treatment), but with three groups and two treatments, a control group, a group with the P1 treatment (Equal Pay), and a group with the Pol treatment (combined policy), made up just of Illinois, since it seems that no state has passed a Pay Transparency law **before** an Equal Pay law. . Furthermore, the Pol treatment includes the P1 treatment ($Pol = P1 \cap P2$). Instead of considering two independent treatments, we can think of what follows as a DiD with several groups and the same treatment of varying intensity (two levels), and we will show that DiD_{P2} is identifiable even if it is not directly observable empirically.

The complete linear model with data collapsed by State would correspond to:

$$y_{s,t} = \alpha + \lambda Post_s + \gamma_1 P1_s + \gamma_2 P2_s + \gamma_3 (P1 * P2)_s + \delta_1 (Post * P1)_{s,t} + \delta_2 (Post * P2)_{s,t} + \delta_3 (Post * P1 * P2)_{s,t} + \mu_t + \sigma_s + X_{s,t} + \epsilon_{s,t} \quad (2)$$

With $Post$ the dummy of the observation of state s after the law ($=1$) or before ($=0$), $P1_t$ that for whether the state is treated ($=1$) or not ($=0$) in period t by the equal pay policy, idem for $P2_t$ but for the transparency pay policy, $(Post * P1)_{s,t}$ and $(Post * P2)_{s,t}$ are the dummies of the interaction effects (i.e. treatment effects) $=1$ if the state is the one treated after the policy change, $=0$ otherwise. μ_t and σ_s are time and state fixed effects, and $X_{s,t}$ is the vector of group control variables, including human capital. All this, of course, with the usual assumption of common trends in terms of the counterfactual.⁸, i.e.:

$$\begin{aligned} E(Y_{t=1,g=1}^0) - E(Y_{t=0,g=1}^0) &= E(Y_{t=1,g=2}^0) - E(Y_{t=0,g=2}^0) \\ &= E(Y_{t=1,g=3}^0) - E(Y_{t=0,g=3}^0) \end{aligned} \quad (3)$$

Nevertheless, let us note the presence of singularities due to perfect multicollinearities in the model, in this case :

- $\gamma_3 (P1 * P2)_s$ is colinear with $\gamma_1 P1_s$ and $\gamma_2 P2_s$
- $\delta_3 (P1 * P2 * Post)_{s,t}$ is colinear with $\delta_1 (P1 * Post)_{s,t}$ et $\delta_2 (P2 * Post)_{s,t}$

The model is not identifiable, so we have to drop terms. Under the assumption of additivity and in a linear framework, we can drop the terms $\gamma_2 (P2)_s$ and $\delta_2 (P2 * Post)_{s,t}$, considering that $\gamma_3 = \gamma_1 + \gamma_2$ (preliminary differences in outcomes between states) and that $\delta_3 = \delta_1 + \delta_2$ (the effect of a combined policy would be the sum of the effects of the independent policies).

We now have an identifiable model, and can decompose it to identify the treatment effect of P2, even though no group has been treated by P2 alone. Let's start by identifying the treatment effects for each policy actually observed, i.e. their DiD estimator. For equal pay policy P1, we can confine ourselves to a simple DiD between the control group and the group treated only by

⁸Let us note that this notation implicitly rules out dynamic treatment effects while assuming that potential outcomes only depend on current treatment, not on past treatments

this policy (for simplicity of notation, the difference between estimator $\hat{\delta}$ and actual value δ is considered implicit):

$$\begin{aligned} DiD_{P1} &= E(Y_{t=2}^{P1}) - E(Y_{t=1}^{P1}) - \left(E(Y_{t=2}^0) - E(Y_{t=1}^0) \right) \\ &= \alpha + \lambda + \gamma_1 + \delta_1 - (\alpha + \gamma_1) - (\alpha + \lambda - \alpha) = \delta_1 \end{aligned} \quad (4)$$

For the effect of the combined policy, we compare Illinois with the control group:

$$\begin{aligned} DiD_{Pol} &= E(Y_{t=2}^{Pol}) - E(Y_{t=1}^{Pol}) - \left(E(Y_{t=2}^0) - E(Y_{t=1}^0) \right) \\ &= \alpha + \lambda + \gamma_3 + \delta_3 - (\alpha + \gamma_3) - (\alpha + \lambda - \alpha) = \delta_3 \\ &= \delta_1 + \delta_2 \end{aligned} \quad (5)$$

So, under the assumption of linearity and additivity (i.e. $Pol = P1 \cap P2 := P1 + P2$, we can identify and isolate the unobserved effect of the P2 pay transparency policy, independently of equal pay, with our two estimable DiD estimators:

$$\begin{aligned} DiD_{P2} &= E\left(TE_{Pol} - TE_{P1}\right) = E(TE_{P2}) \\ DiD_{P2} &= DiD_{Pol} - DiD_{P1} = \delta_3 - \delta_1 = (\delta_1 + \delta_2) - \delta_1 = \delta_2 \end{aligned}$$

However, this hypothesis is debatable. One could imagine non-linear interaction effects between the two policies, giving a stronger result than the simple sum of the two... Or with stronger perverse effects! (For example, in the case where equal pay constraints coupled with pay transparency lead employers to reduce the number of women they hire, thereby reducing the women's work share).

So we've got everything we need... Except for the simultaneity of the treatments. In fact, despite various semantic searches and in various American legal databases, we have not found any state that passed a "simple" Equal Pay Act in the same year as Illinois, in 2003. We will therefore have to use a deferred treatment framework.

5.3 Staggered DiD with several treatments

We have a situation where we have several groups and several treatments (or one treatment with variable intensity) that are "staggered", i.e. the states are not treated at the same time. If we only wanted to study the impact of the Illinois Equal Pay Act of 2003, we could have done a Two Way Fixed-Effect (TWFE) like the reference paper, focusing on Illinois and constructing an artificial state controlled by synthetic control - excluding in particular the states that will pass an Equal Pay Act over the period 2004-2016, otherwise their TWFE would no longer make sense). However, we still want to isolate the pay transparency policy, so we need a control group as well as a group only affected by the equal pay policy.

We could use a staggered DiD, whose model would be :

$$y_{s,t} = \alpha + \delta_1 \text{Treat}_{P1,s,t} + \delta_3 \text{Treat}_{Pol,s,t} + \sum_{t=2001}^{2006} \lambda_t * 1\{year = t\} + \sum_{g=1}^2 \gamma_g * \{group = g\} + \mu_t + \sigma_s + X_{s,t} + \epsilon_{s,t}$$

with $\text{Treat}_{P1,s,t}$ and $\text{Treat}_{Pol,s,t}$ dummies for each state that equals 1 if the state is in the treatment group of P1 or Pol after the policy intervention, and $year$ are a set of time dummies (starting in 2001 and not in 2000 as it is the time trend).

Here, it is sufficient to have at least one state over the period treated by the P1 Equal Pay policy to be able to identify and estimate $DiD_{P2} = \delta_2$, under the counterfactual Common Trends hypothesis and eliminating the effect of the time differential (e.g. exogenous common trend reduction in the wage gap).

Nevertheless, we might wish to have a P1-treated group of consistent size to maximize the quality of the estimate of P1's effect, in order to correctly isolate P2 in Pol. But at the cost of a new implicit assumption! Indeed, staggered TWFE/DiD implicitly make the assumption that the treatment effect is homogeneous, both by period and by treatment. Formally:

$$\forall(g, t), Y_{g,t}(\text{Treated}) - Y_{g,t}(\text{Untreated}) = \text{constant}(\delta_{fe})$$

Consequently, recent literature in theoretical econometrics tends to show that staggered DiDs give generally biased estimators (*inter alia* Sun and Abraham, 2021; Borusyak and Jaravel, 2018; Callaway and Sant'Anna, 2021; **Goodman-Bacon, 2021**; Imai and Kim, 2021; Strezhnev, 2018; Athey and Imbens, 2022; **de Chaisemartin and D'Haultfoeuille, 2020**).

Let us recap: DiD estimates are the unbiased ATE where there is a single treatment period (whatever the treatment effects are static or dynamic). But the adoption of laws or regulations across states or countries are usually staggered, preventing us from using the standard (or TWFE) DiD. Staggered DiD estimates are also unbiased ATE when there is no treatment effect heterogeneity across time or (exclusive "or") treated. But when we combine staggered treatment timing with effect heterogeneity across firms and over time, staggered DiD estimates are really likely to be biased and can even produce the wrong sign of the true average treatment effect (Baker, Larcker and Wang, 2022).

This can lead to biases if the effect is ever heterogeneous over time (for example, if an equal pay law has less impact in 2020 than in 2004 because the margin easily reducible by employers has already been achieved in the meantime due to the "natural trend" or other factors (societal pressure, union action etc.)) or according to the states treated (for instance, the effect of equal pay acts on employment is likely to differ in counties with highly educated female workers and with less educated female workers), c.f. Chaisemartin and D'Haultfoeuille (2020). The impact of the same equal pay policy is unlikely to be the same in Maine as in Alabama.

Thus, $\hat{\delta}_{fe}$ may not estimate a convex combination of treatment effects. It may be desirable to take this heterogeneity into account. Indeed, both the validity of the parallel trends and the treatment effect for each state must be equal to the same constant δ for $\hat{\delta}_{fe} = \delta$. Otherwise, we're really identifying a weighted sum of treatment effects:

$$\hat{\delta}_{fe} = E\left(\sum_{g,t} W_{g,t} TE_{g,t}\right)$$

where the weights $W_{s,t}$ summing to 1 and $TE_{s,t} = ATE$ of group g at time t . See Goodman-Bacon (2019) for the decomposition theorem of the TWFE estimator as a weighted average of all potential DiD estimates, where weights are both based on treatment variance and group size shares. And this can even give us an aggregated average effect... negative, when all the local effects are positive!⁹.

Indeed, this is a bias that Andersson and Ekman (2018) would have encountered if they had attempted to construct a TWFE to estimate the effect of Equal Pay Act laws in staggered treatment, and not simply the impact of the Illinois Equal Pay Act of 2003. A set of alternative estimators for convex combination of heterogeneous and staggered treatment effects have recently been proposed through the "Event Study Design", c.f. *inter alia* de Chaisemartin and D'Haultfoeuille (2020) excluding dynamic effects, and Callaway and Sant'Anna (2021) as well as Sun and Abraham (2021) allowing for dynamic effects. This would give us a particularly fine $\hat{\delta}_{ATT}$ of P1 to subtract from $DiD_{Illinois\ Pol}$ to obtain an equally fine $E(TE_{Illinois\ P2})$.

Finally, let's specify that if we had data from 2000 to 2022 with the table of the various Equal Pay state laws as well as those combining it with Pay transparency (a significant wave of states passed Equal Pay Act laws in 2019) and if we were interested not only in the impact of the 2003

⁹In fact, if we simplify to 2 periods ($t = pre$ or $post$) with K the states treated before and L the states treated lately, formally the DiD estimator is equal to (indexes are dropped for convenience):

$$\begin{aligned} DiD &= E(Y|t = post, state = L) - E(Y|t = pre, state = L) \\ &\quad - (E(Y|t = post, state = K) - E(Y|t = pre, state = K)) \end{aligned} \quad (6)$$

$$= E(Y^1 - Y^0|t = post, state = L) + E(Y^1 - Y^0|t = pre, state = K) - E(Y^1 - Y^0|t = post, state = K)$$

under the assumption of parallel trends on Pol^0 i.e. on

$$\begin{aligned} &E(Pol^0|t = post, state = L) - E(Y^0|t = pre, state = L) \\ &\quad - (E(Y^0|t = post, state = K) - E(Y^0|t = pre, state = K)) \end{aligned} \quad (7)$$

If the treatment effect is stable over time (i.e. does not depend on t) or if it is homogeneous (i.e. does not vary between states), then DiD is identified.

Indeed, if it is stable over time :

$$E(Y^1 - Y^0|t = pre, state = K) - E(Y^1 - Y^0|t = post, state = K) = 0$$

And so $DiD = E(Y^0 - Y^1|t = post, state = L)$ Similarly, if it is stable over time :

$$DiD = E(Y^1 - Y^0|t = post, state = K)$$

However, if the treatment varies according to state and/or time, DiD is the sum of three treatment effects:

$$\begin{aligned} DiD &= E(P1^1 - P1^0|t = post, state = L) + E(P1^1 - P1^0|t = pre, state = K) \\ &\quad - E(P1^1 - P1^0|t = post, state = K) \end{aligned} \quad (8)$$

one of which has a negative weight (-1). The estimator can therefore be artificially negative while the three treatment effects are strictly positive (violation of "no sign reversal" property). DiD is then no longer identified.

This result is generalized by de Chaisemartin and D'Haultfoeuille (2020). They propose an alternative estimator that allows for heterogeneous treatment effects as long as the trends are parallel, by checking for the presence of negative weights (using the *did_multiplegt* package).

pay transparency policy in Illinois, but **more broadly in the impact of pay transparency in general in the United States**, we could identify and estimate an approximation of the average effect on the states (ATE of P1 and Pol, and therefore of P2 by subtraction and assumption of linearity and additivity), through an **Event Study Design with multiple treatments** such as most recently developed by Sun and Abraham (2021) with mutually exclusive treatments, and de Chaisemartin and D’Haultfoeuille (2022) with potentially non-exclusive treatments, which is our case here. An example of a model for us would be:

This would be the "academic research" equivalent of the present assignment if we wished to do a real research paper on the impact of pay transparency measures in isolation from comminatory equal pay measures.

6 DiD between Vermont and Illinois on Pay Transparency

All these theoretical discussions on the empirical design to be used can only succeed if we have, as we have demonstrated, at least one state that was treated by the Equal Pay Act policy alone between 2001 and 2006 (2001 to have the year 2000 in order to test the pre-trends) and therefore the hypothesis of common trends with Illinois is plausible, or else a state treated only by a Pay Transparency Act. After a systematic semantic search (e.g. "Equal pay act 2002", "Equal pay act 2005", "Equal pay state law 200X" etc.) and an in-depth search of various academic publication databases and US institutional sites, we found that to our knowledge only Vermont passed a Pay Transparency Act in 2005 (wage disclosure only, for both private and public sector). There have been several occurrences post-2008 (Colorado, Maine) and especially in the 2010s, with a major wave in 2019. Every legislature has also seen an attempt to pass the Fairness Pay Act at federal level, which would oblige companies to provide pay transparency, but this has been rejected every time. Nevertheless, we found one state treated only by policy P2, without being treated by policy P1 (Equal Pay Act) - Vermont won't pass an Equal Pay Act until 2013. Nevertheless, we have not been able to identify any state over the period 2000-2005 that has only passed an Equal Pay Act without Pay Transparency.

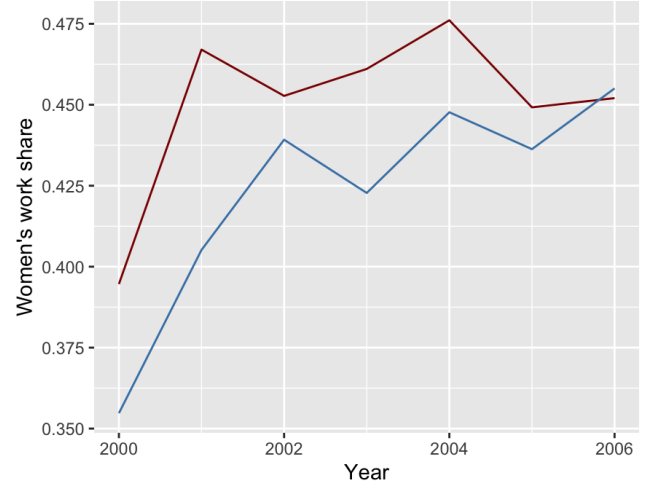
We therefore have: one state treated simultaneously by P1 and P2, i.e. by Pol (Illinois 2003), and one state treated only by P2 (Vermont 2005). Under the assumptions of linearity and additivity, we can therefore identify the isolated effect of the P2 Pay transparency policy.

We can therefore build an event design study to identify the effect of pay transparency in Illinois by subtracting from the total effect of the Equal Pay Act of 2003 and applied in 2004 the effect of pay transparency without equal pay act in Vermont, voted in 2005 and applied in 2006 - event study design to take into account the staggered character (Pol in 2004, P2 in 2005).

A first element to check is the parallelism of pre-trends, to reinforce the common trends hypothesis. We visually examine the evolution of the GWG of the two States between 2000 and 2003:



Time trends of Wage Gap between female and male employees in Illinois (blue line) and in Vermont (red line).



Time trends of women's labour market participation for Illinois (blue line) and average for Vermont (red line).

The pre-trends do not appear to be parallel, particularly for GWG, which is our main variable of interest. Furthermore, we find that there is no reduction in the GWG in Vermont between 2005 and 2006 - it is more likely that the data stops before the effect of implementation in early 2006 and there was no anticipatory effect by employers, rather than Pay Transparency Act passed in 2005 having no effect.

We can reinforce this graphical approach by testing for a *placebo* treatment: we carry out a DiD over the period 2000-2006, integrating all leads and lags (4 pre-treatment periods and 3 post-treatment period - 4 because there seems to be no effect of announcement of the law and therefore of anticipation by Illinois companies in 2003, which might have sought to reduce the gender pay gap ahead of the effective application in 2004):

$$y_{s,t} = \sigma_s + \lambda_t + X_{s,t} + \sum_{\tau=1}^3 \beta_{t-\tau} D_{s,t} + \sum_{\tau=1}^3 \beta_{t+\tau} D_{s,t} + \epsilon_{s,t}$$

$$\Longleftrightarrow$$

$$y_{s,t} = \sigma_s + \lambda_t + X_{s,t} + \beta_{-3} D_{s,t} + \beta_{-2} D_{s,t} + \beta_{-1} D_{s,t} + \beta_1 D_{s,t} + \beta_2 D_{s,t} + \beta_3 D_{s,t} + \epsilon_{s,t}$$

Here, we include the interactions of the time dummies and the treatment indicator for the first three pre-treatment periods and we leave out the one interaction for the last pre-treatment period due to the dummy variable trap (perfect multicollinearity). If the outcome trends between treatment and control group are the same, then β_{-3} , β_{-2} and β_{-1} étant alors censés être nuls ou a minima inférieurs à une marge "acceptable" en l'absence d'un quelconque traitement aux seuils de significativité statistique habituels, i.e. the difference in differences should not be significantly different between the two groups in the pre-treatment period¹⁰.

We find that these β_{-t} are non-zero and significant (tables available on request). This confirms the rejection of pre-trend parallelism.

¹⁰Moreover, β_j with $j > 0$ may not be identical. For example, the effect of the treatment could accumulate over time, so that β_j increases in j

The absence of pre-treatment parallelism does not necessarily imply an absence of counterfactual parallelism of the treated if untreated group with the control group in the post-treatment period. Nevertheless, it seems perilous to carry out a TWFE / event design study under such conditions. The risk of identifying a biased Pay Transparency ATE in Illinois seems too high.

7 Finally, back to a trivial DiD: estimating the cumulative effect of the 2003 Act

This impossibility loses much of the interest of the present work¹¹ as to the initial research question which was *"Does the pay transparency act affect gender equality in the labour market?"*, but we can still take it at face value and study not the impact of pay transparency, but of the pay transparency act of 2003 in Illinois, synonymous with the simultaneous and joint 2003 Equal Pay Act. We can identify and estimate the effect of this combined policy, we just won't be able to identify and estimate the isolated and separate impact of pay transparency in relation to the equal pay obligation, but only the interaction effect of the two measures.

We have two possibilities:

- either perform a Difference-in-Difference-in-Difference (DDD) at the level of individuals (or households) with a Mincer's equation and dummies for gender and for whether they are in a treated state or not.
- either collapsing the data by state, taking the median - or mean to control - value of each state (and potentially constructing an "artificial control" state through Synthetic Control Method) if building a control group with common trends with Illinois is not viable);

Normally, the two estimators (the second weighted by the number of observations in the cell) should give the same results, the only difference being the degree of aggregation. Nevertheless, the problem with aggregation at state level is the collapse of statistical power, and hence the risk of type II error.¹² Causal inference requires sufficient sample sizes, especially as in panel data (which is the case when collapsing cross-sectional data by State), power depends on the correlation structure between repeated measures of the same unit (here the State), typically positively correlated, which is detrimental to power estimation as well as statistical analysis. If general linear models (GLM) for statistical analysis and power estimation are well known (Self and Mauritsen, 1988) for randomized data, it is not really the case for non-randomized studies using DiD analysis. We would have significance and power problems. The use of Synthetic Control Method to construct a control entity is an alternative (we could also include within state variation in the aggregated regression to reduce standard errors).

¹¹Loss of interest because the impact of equal pay laws has already been studied (c.f. Literature review)

¹²The treatment group is of size 1 or 2. The control group is smaller than 50 states, in order to select only those states that make the common trends hypothesis realistic. Yet, the statistical power of a test, i.e. the probability of rejecting the null hypothesis (here the hypothesis that the 2003 law had no impact on the variables of interest) equivalent to $1 - \beta$ with β the risk of false-negative, depends on the type I risk (the famous threshold α), the effect size... and sample size (relative to variance). For a two-sided two-sample t-test with a significance level of 5%, a control group of size 49, an alternative hypothesis "two.sided" and an effect size corresponding to that found by Andersson and Ekman (2018), the use of the function *pwr.t2n.test* to find the necessary size of the treated group in order to have a power of 80% (package *pwr*) is not even possible because the effect size is too small.

In the case of disaggregation to the individual level, (while controlling for household composition), we could mobilize a Mincer-style wage/human capital equation, with dummies for gender, instead of doing a log-difference/ratio, to do a Difference-in-Difference-in-Difference (see Imbens and Wooldridge, 2007). For example, following Kim (2015), the DDD specification could be:

$$\ln(w_{i,s,t}) = \beta_1 Women_{i,s,t} + \lambda T_t + \sigma S_s + \beta_2 (Women_{i,s,t} * PayTr_{s,t}) + \beta_3 (Women_{i,s,t} * T_t) + \beta_4 (Women_{i,s,t} * S_s + \beta_5 (T_t * S_s) + \beta_6 X_{i,s,t} + \epsilon_{i,s,t} \quad (9)$$

The dependent variable is the natural log of the real hourly wage for each individual i in state s and time t . $Women_{i,s,t}$ is the gender dummy, the effect of underpaying female workers controlling for others factors ($X_{i,s,t}$), S_s is a matrix of state effects, T_t a matrix of fixed effects by year, $Women_{i,s,t} * S_s$ captures political climates toward women or other state laws that may affect the pay for women and the gender pay gap across states, $PayTr_{s,t}$ indicates those living in a state s in which pay secrecy laws were in effect in year t , $Women_{i,s,t} * PayTr_{s,t}$ measure the effect of outlawing pay secrecy on women's wages in particular (so β_2 is the coefficient of interest), ($Women_{i,s,t} * T_t$ controls for trends particular to women, such as the general tendency for the wage gap to decline over time, and $T_t * S_s$ controls for unobservable factors by state-year cells. We would then need to cluster at state level to obtain robust standard errors (which would mechanically be larger). Finally, thanks to the household ID in the available data, we could directly study GWG trends at household level, but this is another research topic.

Nevertheless, we'll continue here with the logic of aggregation by state, with the construction of a synthetic control state by SCM. Similarly, including co-variates (control variables) at the individual level in the aggregate regression ($X_{i,s,t}$ instead of $X_{s,t}$) enables us to obtain the same estimates while reducing standard errors. Unfortunately, since the variable of interest is a log difference (or ratio) between two categories (men and women), it does not seem feasible to include co-variates at the individual level.

7.1 Model

In this context, we can perform a DiD between the treated Illinois and the rest of the USA as a control group, with time and state fixed effects and collapsed by State (we will not need to cluster). Our specification will be:

$$y_{s,t} = \sigma_s + \lambda_t + \gamma Post_s + \delta Illinois_t + \beta (Post * Illinois)_{s,t} + X_{s,t} + \epsilon_{s,t}$$

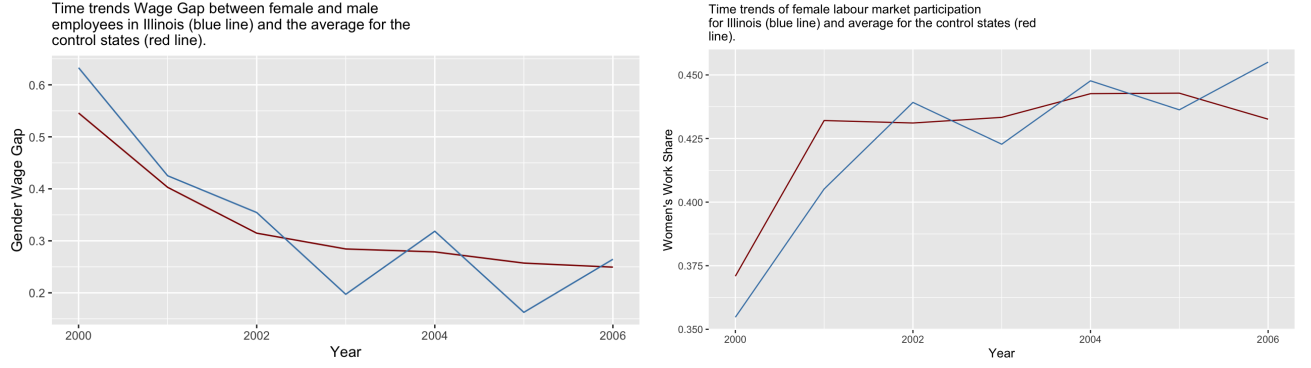
where $(Post * Illinois)_{s,t}$ is the interaction term for being treated (equivalent to being Illinois) and for being in the post-treatment period. So, β is our coefficient of interest, the DiD estimate (which is the ATT if not biased). σ_s and λ_t are the state and time specific fixed effects, and $X_{s,t}$ the vector of control variables. We want to study the evolution of the "residual" GWG, i.e. the GWG not explained by gender differences in human capital. We therefore want to control for the said capital. The standard literature for this purpose controls for both *Education* (with years of schooling and categories of academic degree) and *Work Experience* (while differentiating

between experience in full-time work, which is associated with positive returns, and experience in part-time work, which is indicative of low earnings and hence negative returns (Gornick and Jacobs, 1996; Olsen and Walby, 2004)). Unfortunately, the available census data do not include work experience. We will therefore control only for the state median of the variable *sch*, "Educational attainment" (integrating Bachelor and Advanced degree) - and not *Educ99* which includes more than 50,000 NA, leaving open the risk of truncated sample, restricting identification. We also control for the median age of each state, reducing the impact of extreme values, as well as the state averages of individual ethnic indicator dummies.¹³, of the number of minor children per individual (through household ID and relate variable) and the fact of being married, a method making it possible to have the proportion of the group concerned by these variables in each state at each period (e.g. for a state with N the number of statistical individuals in the state: $\frac{1}{N} \sum_{i=1}^N \mu_i$ with μ_i the dummy for the individual i). As mentioned *supra*, we cannot control for individual variables given the outcome of interest (which does not allow us to reduce SE).

To determine whether it is preferable to use random or fixed time and State effects, we perform a Hausman test, where the null hypothesis is that the errors are random (the random effect model being the best), while the alternative is that they are correlated with the regressors (the fixed effect model being the best). The null hypothesis is rejected at all significance levels. This is necessary because collapsing by state leads to a panel data situation rather than a cross-sectional one, running the usual risks (c.f. Bertrand, Duflo and Mullainathan, 2004). Thus, under the assumption of common trends, the DiD β estimator identifies the ATT i.e. the treatment effect of the 2003 Act on Illinois over the 2004-2006 period (or 2003 if we want to test for a possible anticipation effect). We will run it on the year 2004, then 2005, and 2006, to obtain the annual impacts of the Act, in a logic equivalent to a TWFE but manually.

But obviously, as we have seen with Vermont alone, the Common Pre-Trends hypothesis for all these states is implausible. We can try to strengthen the control group by removing states from which differences in trends can be assumed: we remove states that are too different, such as the states that already had at the time an equal pay law (Indiana, California, Michigan), states that are structurally specific (Hawaii, Alaska, and the District of Columbia), states with an exceptionally large GWG or increasing over the period (Virginia, West Virginia, Vermont), or exceptionally low (New Jersey) and the state which in 2023 has still not passed an Equal Pay Law (Mississippi), both illustrating strong social and political conservatism. We obtain a control group of 39 states.

¹³"White", "Black", "Hispanic", "Other race". We drop the "Other race" category in our mean control covariates to avoid multicollinearity.



Building a robust control group of Common Pre-Trends states is likely to prove time-consuming, with 49 more DC states to study¹⁴. It might be possible to run a set of placebo pre-trends tests on states that have passed similar laws following Illinois, probably close to its trends¹⁵. Nevertheless, we'll use the Synthetic Control Method (SCM) to build an artificial "Illinois" control. Compared to Matching Methods, SCM not only matches on covariates (i.e. pre-treatment variables), but also outcomes. Thus for causal inference you do not need to have similar covariates in the control and the treated groups (c.f. Abadie, 2021 for a review). It maximizes the observable similarity between control and treatment and can be used in cases where no untreated case are similar on matching dimensions with treated cases.

Here, we build a control group where the common trend assumption is likely to hold by combining weighted averages of the outcome predictors for the states which best mimic the values of Illinois. To build this donor pools, here too, we remove states that are too different, such as the states that already had at the time an equal pay law (Indiana, California, Michigan), states that are structurally specific (Hawaii, Alaska, and the District of Columbia), states with an exceptionally large GWG or increasing over the period (Virginia, West Virginia, Vermont), or exceptionally low (New Jersey) and the state which in 2023 has still not passed an Equal Pay Law (Mississippi), both illustrating strong social and political conservatism.

The "Synthetic control state" is built such that the weight W minimize the sum of the quadratic difference between the predictors in the treated state (Illinois) and the same predictor in the synthetic state of Illinois (to have the smallest possible differences between Illinois and synthetic Illinois prior to the intervention). The DiD naturally follows by subtracting the sum of the values of the control states weighted by the weight identified above from the value of the Illinois outcome treated. It gives the ATT (average treatment on the treated) with a synthetic observable counterfactual Illinois - a reasonable comparison state for which problems with parallel trend assumptions can be seemed as resolved. The difference between these two will be interpreted as the effect of the policy change. We apply the same method to the Women's work share, our other outcome of interest.

¹⁴To be rigorous, parallel pre-trends should be accompanied by the "common shocks" hypothesis, i.e. in post-treatment, exogenous forces affect both control and treated groups, and equally (see Ryan, Burgess, and Dimick 2015), which in this case is not testable (perhaps with unit roots tests on the stationarity of the series after 2004?). It would need a deep study of those states' recent history

¹⁵Finally, the positivity assumption is trivially respected at the state level, i.e. the fact of passing such a law is not entirely determined by the control variables collapsed by State, even if we can obviously assume a certain endogeneity (e.g. a State with a highly educated population may possibly be more progressive? We could also imagine controlling for historical Republican/Democrat political color). Formally: $\forall X, 0 < P(Treat = 1|X) < 1$.

8 Results

8.1 Fixed Effects OLS Results

Table 1 presents the DiD estimates following the Act of 2003 in Illinois, for the years 2004 (implementation), 2005 and 2006, for both raw GWG, GWG controlled for our covariates specified above, and women’s work share on labor market. Each “DiD” estimate therefore corresponds to the effect of the Act in the year in question. There appears to be no impact on GWG in 2004, the year of implementation, with a simple drop of 1%, and no negative effect on women’s work share. It’s in 2005 that the Act seems to materialize its effect, with a significant reduction in GWG (10%), with no negative effect on women’s work share either (despite an overall drop in this share in Illinois, but apparently not linked to the Act). Finally, 2006 simply saw a 2% reduction in GWG, as the Act seemed to have exhausted its transformative capacity. Moreover, it even saw a 4% increase in the women’s work share linked to the Act. **But none of these results is significant at any of the usual thresholds** (although the inclusion of covariates reduces the standard errors, but not enough - the impossibility of including the *Work experience* also prevents confidence intervals from being reduced). This suggests (modulo statistical significance) a non-negligible positive effect of the Equal Pay and Pay Transparency Act of 2003 on the Illinois GWG, with no negative perverse effect on women’s work share.

Table 2 presents estimates not just for private-sector salaried workers, but for all salaried workers, including those in the public sector (federal and local), as a robustness check. The standard errors decrease significantly, but here again none of the results for the impact of the Act are significant. Note, however, that the estimates are lower, which may be interpreted as meaning that the GWG of public sector employees was not impacted by the Act, a realistic possibility as public sectors may have had prior regulations prohibiting wage discrimination, and/or are more inclined not to discriminate on the basis of taste based, and/or comparisons between equivalent positions are far more numerous. Results are similar for the SCM.

It is not recommended to carry out DiDs over long periods, due to the fragility of the common trends hypothesis. Here, the 3-year period provides a good framework, showing that it does take one year for the Act to take effect. Furthermore, checking for anticipatory effects (i.e. employers who, on seeing the bill discussed and voted on in 2003, would seek to regularize, and therefore reduce their GWG, and/or reduce the hiring of women, even before the law was implemented, is not useful given the graphical analysis showing the GWG increasing in 2003. We nevertheless carry out a placebo test for 2003 to verify this. The estimates are almost nil and non-significant.

The problem here, as explained above, is the improbability of the common trends hypothesis between the control group and Illinois, and even in terms of pre-trends, limiting the possibility of inferring the effect of the Act and undermining statistical significance. Let us compare with the results of the estimates made with the Synthetic Illinois.

Table 1: OLS model results

	Dependent variable:										
	GWG (2004)		Work share (2004)		GWG (2005)		Work share (2005)		GWG (2006)		Work share (2006)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)		
Constant	0.384*** (0.010)	0.334 (0.374)	0.442*** (0.120)	0.384*** (0.010)	0.842** (0.394)	0.329** (0.127)	0.384*** (0.011)	0.437 (0.418)	0.417*** (0.127)		
Illinois	0.018 (0.065)	0.045 (0.047)	-0.015 (0.015)	0.018 (0.065)	0.037 (0.049)	-0.014 (0.016)	0.018 (0.067)	0.041 (0.050)	-0.015 (0.015)		
Post	-0.106*** (0.023)	-0.003 (0.019)	0.004 (0.006)	-0.126*** (0.023)	-0.021 (0.020)	0.004 (0.007)	-0.136*** (0.024)	-0.006 (0.021)	-0.015** (0.007)		
DiD	0.022 (0.144)	-0.013 (0.104)	0.034 (0.033)	-0.114 (0.146)	-0.127 (0.110)	0.008 (0.035)	-0.002 (0.149)	0.0002 (0.113)	0.042 (0.034)		
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes		
Observations	195	195	195	195	195	195	195	195	195		
R ²	0.100	0.553	0.482	0.142	0.537	0.432	0.149	0.544	0.451		

*p<0.1; **p<0.05; ***p<0.01

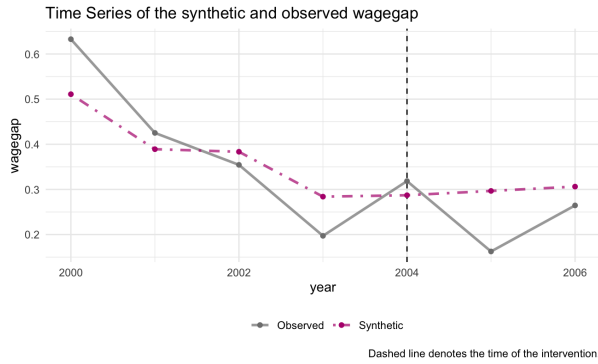
Table 2: OLS model results

Dependent variable:												
	GWG (2004)		Work share (2004)		GWG (2005)		Work share (2005)		GWG (2006)		Work share (2006)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)			
Constant	0.352*** (0.009)	0.731** (0.344)	0.379*** (0.108)	0.352*** (0.009)	0.981*** (0.345)	0.303*** (0.111)	0.352*** (0.009)	0.908** (0.369)	0.352*** (0.111)	0.352*** (0.111)		
Illinois	0.038 (0.057)	0.061 (0.042)	-0.019 (0.013)	0.038 (0.057)	0.058 (0.042)	-0.018 (0.014)	0.038 (0.058)	0.056 (0.043)	-0.019 (0.013)			
Post	-0.087*** (0.021)	0.004 (0.017)	0.006 (0.005)	-0.114*** (0.020)	-0.021 (0.017)	0.003 (0.006)	-0.122*** (0.021)	-0.010 (0.019)	-0.014** (0.006)			
DiD	0.019 (0.128)	-0.001 (0.092)	0.025 (0.029)	-0.087 (0.128)	-0.102 (0.094)	0.010 (0.030)	0.032 (0.129)	0.028 (0.097)	0.036 (0.029)			
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes		
Observations	195	195	195	195	195	195	195	195	195	195		
R ²	0.090	0.548	0.530	0.150	0.560	0.479	0.158	0.552	0.497	0.497		

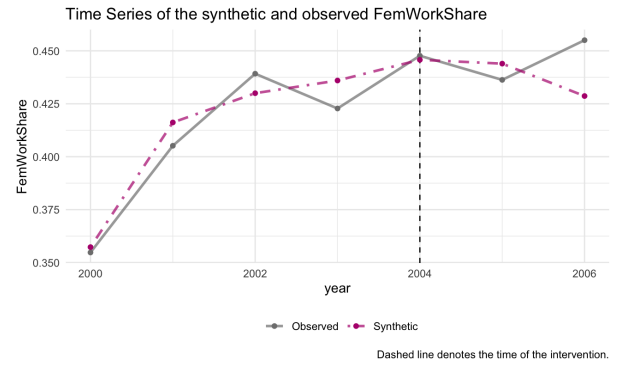
*p<0.1; **p<0.05; ***p<0.01

Note:

8.2 SMC results



Difference in median hourly real wages between men and women in Illinois and synthetic Illinois.



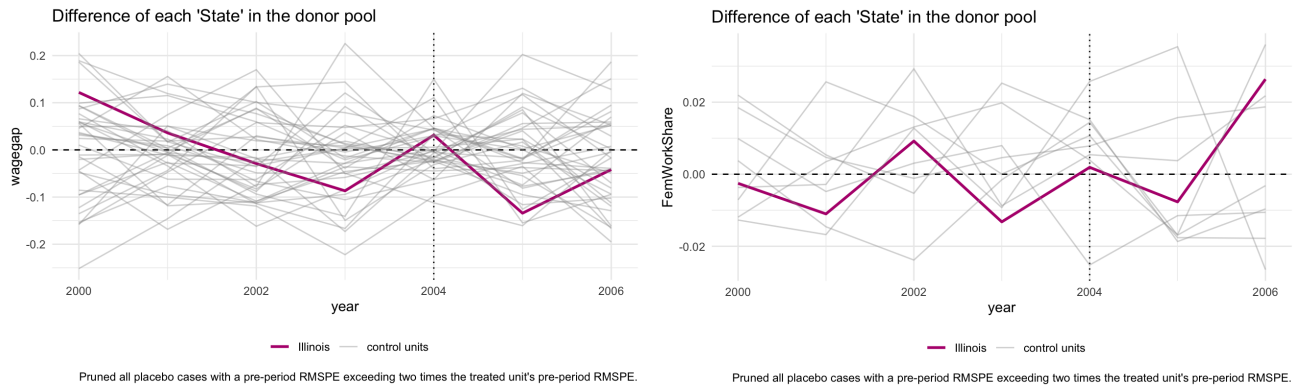
Difference in median share of hours worked by women employees as a share of the total hours worked in labour market in Illinois and synthetic Illinois..

Regarding the SCM, we can see quite a good fit between factual and counterfactual Illinois in the pre-treatment period for both the GWG and women's work share. But the trajectories differ after the Act for the GWG: there's a clear drop in GWG in treated Illinois in 2004, before a reaugmentation after 2005. It is less diverging for women's work share, although there has been a slight decline following the Act, before a reaugmentation here to. Let us note that the counterfactual Illinois registers a decline of this share¹⁶.

Thus, we were able to estimate by Fixed Effects difference-in-difference a relatively significant impact of the 2003 Act on the reduction of the Illinois GWG, with no negative impact on the women's work share. However, these results are not statistically significant. Similarly, to get around the problem of the common trends hypothesis, we have used the Synthetic Control method, and we do indeed observe a fall in GWG in the treated state compared with the synthetic counterfactual state. We are now carrying out a robustness test using placebo SCM to ensure that this effect is not random, but due to the Act.

¹⁶The weights tables Weighting tables for states and predictors are available on request.

8.3 Placebo results



Placebo test for GWG. The dark line represents the difference between Illinois and synthetic Illinois. Grey ones represent the corresponding values for the donor states.

Placebo test Women's work share. The dark line represents the difference between Illinois and synthetic Illinois. Grey ones represent the corresponding values for the donor states.

These graphs represent the difference in our two variables of interest in values between an untreated state (observable) and the same synthetic treated state (hypothetical), with Illinois in the opposite situation for the darker curve. Only two of the 38 states in the control group recorded a greater placebo drop in GWG than Illinois, reinforcing the idea that GWG variation in Illinois is not random - even if there is a wide dispersion on the graph, which nuances this result. As for the women's work share, the situation is much less clear-cut, with several states having greater positive or negative variations than Illinois. It is less possible to infer a causal relationship.

On the other hand, if we look at the p-values of the SCM ATT computed by calculating the root mean square prediction error (RSMPE) method, based on the placebo test, none are below the usual thresholds. As for the p-values of the estimated DiD, depending on standard errors, the situation is identical (c.f. Table 1). It is therefore not possible to reject the hypothesis that the evolution of both the GWG and the women's work share in Illinois is due to chance, and not to the 2003 Act.

9 Discussion

Despite the estimates seeming to show a non-trivial impact of the 2003 Act on GWG reduction without having a negative effect on women's work share, they are not significant at any level, neither for the FE-OLS DiD nor for the SCM (even if the visual analysis of the SCM seems to give a slight significance of GWG reduction for Illinois). There are several possible reasons for this lack of significance. Firstly, the fact that we were unable to control for all the covariates of interest (notably *Work Experience*, limiting the reduction in standard errors).

Secondly, if discrimination against women on the job market is not *taste-based*¹⁷ but statistical i.e. linked to differences in choice according to gender and gendered preferences (for women,

¹⁷In reality, even those taste based are not based on tastes but on "prejudice," "belief," and "stereotype", (Yasuda, 2023).

choice of part-time work, acceptance of lower pay in exchange for flexible working hours, in particular to take on care responsibilities etc.) then such a policy has no significant impact, except potentially for positions with equal skills, choices and human capital. Nevertheless, it is consistent econometric literature that the USA is affected by *taste based* discrimination (Yasuda, 2023).

Thirdly, there may be a lack of awareness of the law, both for employers and employees (which affects bargaining power). This depends on the communication and control mechanisms put in place by the Illinois authorities.¹⁸

Fourthly, and relatedly, it depends on the content of the law, the intensity of the measure and its repressive potential. For example, for the Pay Transparency part of the Equal Pay Act, as explained by the ILO (2022), it can take a wide variety of legislative and regulatory forms, from allowing employees to request and access information on pay levels in their enterprise to requiring employers to disclose individual pay information to employees, as well as prohibiting employers from requesting an employee's or prospective employee's salary history, creating an independent body to provide employers with equal pay certification if they meet certain requirements around gender-neutral pay, and even obliging enterprises with a certain threshold level of employees (for example, 50) to publish information on gender and pay within their organization. The pay transparency introduced by Illinois is therefore minimal: it simply implies a prohibition on discrimination/sanctioning the research and disclosure of salaries by employees, and can therefore fall within the scope of theories of discrimination in hiring but also in promotion, and in the theory of salary negotiation. Swiss and Danish laws have had more significant results, potentially also because they were more intense and repressive.

Fifthly, initiating legal proceedings against an employer in the event of a breach of the 2003 Act is a heavy burden for a female employee, in terms of psychological, emotional, financial and even social costs, and the risk of backfire within the company or on the job market in the event of a signalling effect. This can reduce the number of appeals, and the incentive for the employer.

Sixthly and finally, this also requires the existence of equivalent and comparable positions, with equal human capital, in the same company, which is not necessarily the case, especially in sectors that are segmented according to gender from the moment of hiring, and due to the non-random selection of women in employment (c.f. *infra*). This applies not only to women employees, but also to employers acting in good faith, who will not necessarily have the comparative elements needed to identify the "fair wage".

As for the women's work share, which is also not significant either in the FE-OLS DiD or in the SCM, the values of the estimates are in any case very close to zero. This suggests that the 2003 Act had no negative effect on this variable. Either because it had no impact at all, including on the GWG, meaning that there is no perverse effect to a non-existent effect. Or, if the Act did have a real impact, it had no perverse effect, suggesting that women were then paid

¹⁸ According to a Danish National Centre for Social Research study, only about 30 per cent of the 740 enterprises surveyed compiled gender-specific pay statistics or produced pay transparency reports. About 50 per cent of non-compliant enterprises said that they were not aware of their obligations; while another 25 per cent were unaware that they were in fact included in the set of enterprises that should report on pay. Most survey respondents indicated that there was no gender pay gap in their enterprise. The enterprises that complied and provided reports mainly limited the information provided to the minimum legal requirements. The definitions of pay used by the enterprises varied considerably by the wage period (hourly wage, daily wage, and so on), inclusion or not of absenteeism, and other factors., ILO (2022).

below their marginal productivity (undervalued on the labor market).

Incidentally, there are other methods of estimation, SCM placebo tests, ways of calculating p-values, etc., but as file drawer effects and publication bias are already so high (Altman, 2004; Camerer et al. 2016; Weintraud, 2016), we are not looking to absolutely find a significant effect. Nevertheless, it's worth remembering that significance testing comes with frequent misuses (especially when p-values became standard measures of statistical evidence). Ziliak and McCloskey (1996, 2004, 2008) defend the idea that statistical significance levels should not be used as indicators of actual relevance for economic policy decisions, as the p-value is not an indicator of effect size by itself. According to them, this confusion brought a lot of damages to economics and society (in terms of policy making). Thus, an absence of statistical significance should not be a brake on the implementation of anti-discrimination policies, but should on the contrary be a driving force for research, including qualitative and sociological research, to pursue this goal.

What's more, as we'll discuss later, we've only tested for the GWG median and state average, we haven't looked at the impact of the Act within the wage distribution. It is entirely possible that the GWG of low-wage earners has been significantly reduced, as was the case in England (Dex et al., 2000). In addition, we need to study the impact of controlling for the level of unionization. In England, the effect of the Equal Pay Act has been strongly mediated by union action (Deakin et al., 2015), as unions are informed of these laws, help with procedures, and participate in the balance of power (in addition to building broader wage bargaining capacity).

It should also be pointed out that, while the sampling for Illinois appears to be a large, independent and representative cross-sectional repeated sample and can therefore lead to results with external validity for the Illinois labor market, its scope for the other American states seems unrealistic, given the many socio-cultural, legislative, political and economic differences between them. This calls for further research (Event Study Design over the period 2000-2020 with staggered treatment, for example).

Finally terms of design, had we had a more precise Zip code including the cities of the statistical individuals, we could have had a "DiDD", Difference in Discontinuity Design type approach (c.f. *inter alia* Ludwig and Miller, 2007, Calonico et al., 2014, Grembi et al., 2016). On a city cut out Illinois with another state or on two "glued" cities separated by the border, one could justify the assumption of similarity between the treated and untreated group, with the border as the only discontinuity and perfectly arbitrary (even if it means controlling for the few differences in terms of labor laws etc). It should also be remembered that, although presented pedagogically primarily as a DiD, the work of Card and Krueger (1994) is also accompanied by a matching across the border and a spatial discontinuity approach in restaurants close to the NJ/Penn border. Let's now turn to a more theoretical reflection on methodological issues concerning the theoretical constructs used, especially the GWG and women's work share.

9.1 Non-random selection

Beyond the many limitations and challenges presented in the theoretical section regarding empirical design, estimation methods and the extent of available data, there is also very likely to be another source of bias in our own estimation of the GWG: the possible existence of non-random selection into employment. Indeed, we can imagine that a certain number of women refuse to

enter the workforce, or have withdrawn from it, and that if they were paid, they would indeed have a lower wage than men - or even that this could be a cause of their non-activity. Their wages are therefore not observed, but if they were, and if they were indeed lower than men's for the same task, then the GWG would be higher. Bryson et al (2020) correct their net GWG for non-random selection on British cohort data and show that otherwise the net GWG is underestimated (even when controlling for human capital: it is around 4 percentage points larger at age 23 when selection-adjusted. Similarly, the adjusted GWG is around 3 percentage points larger at age 30 than the GWG between employees. However, adjusting for selection into employment makes no difference to the size of the GWG later in life (after 40).

One possibility would be to create a potential GWG to take account of this non-random selection in employment, using Propensity Score Matching. We can adopt an approach similar to Neuburger et al (2011) where they adjust women's and men's wages by imputing a wage for four types of individual: those in employment without a wage observation; the unemployed; the self-employed; and the economically inactive. These "potential" (imputed) wages would come from nearest-neighbour wage "donors" defined as those in the waged employment group at the same census from the same sex who are nearest in their propensity for waged employment to the non-waged individual. The nearest neighbours would be identified through propensity score matching where the propensity for waged employment is estimated for each individual each year. The probits for the (0,1) being in waged employment at the time of the census would be run separately for men and women so that nearest neighbours who are "donors" of their wage to the non-waged are drawn from the same sex. We could use covariates to match those with and without wages, and enforce common support by dropping cases whose propensity for waged employment falls below the lowest probability for the waged employee sample at that census.

9.2 On the endogeneity and measurement issues of human capital

As mentioned above, traditional neo-classical theories explain GWG primarily in terms of differences in human capital between men and women, and gendered preferences and choices. We have shown that a non-trivial part of GWG is not explained by human capital, demonstrating the presence of discrimination and other factors. But this approach is itself open to criticism. The scientific literature also criticizes the methodology and results of the human capital applied to the GWG analysis. Already, from a conceptual point of view, human capital is based on the axioms and paradigms of rational choice theory, extricating itself from endogeneity with social structure. In the same way that, in Oaxaca-Binder-type decompositions (Blinder, 1973; Oaxaca, 1973), wage inequalities are separated into a share explained by the difference in skills and an "unexplained" share referring to discrimination/choices and preferences and caetera¹⁹ - and epistemologically serve to "demonstrate" the existence of discrimination. Yet this essentializes, precisely "ontologizes" the difference explained, as if the difference in human capital were naturalized and not contingent and determined by social dynamics (self-censorship and social norms, differentiated and unequal access to education, differences in guidance advice, the constraint of early motherhood, racism at the time of the first studies using this decomposition etc...), which themselves may come under discrimination (taste-based and intentional or statistical, uncon-

¹⁹Kunze (2008), Lips, (2013), Ponthieux & Meurs (2015) for such GWG decompositions.

scious, structural...). See Brochier (2020) for a formal critique of this decomposition. Here, human capital may itself be determined by socio-institutional causes and not just by rational choices and socially-neutral variables à la Becker (1962) or Mincer's (1974) income function - ditto for the non-random selection of women in employment; even if non-observable preferences may also play a part in this selection, such as type of activity - in the knowledge that these preferences are themselves partly endogenous to already existing segmentations due to hiring discrimination or sexist professional universes.

There is also a second criticism of human capital evaluation methods (Rotman & Mandel, 2022). The authors here have avoided this bias by studying absolute values, but when human capital is mobilized, it is often to compare rates of return, for example of personal investment in education, on potential earnings (e.g. Diprete & Buchmann, 2006). Yet, often, even though the rate of return may be higher for women, the premium in absolute terms is usually strictly lower than for men (Ibid.). They do not inform us in terms of gender differences in real value concerning these investments and choices (Pekkarinen, 2012). Even more trivially, controlling for human capital as is done in the article implies comparing a male-female counterfactual with the same human capital i.e. the same coefficients for the covariates... And therefore an equal return for this human capital even though its return is different according to gender for the same quantity, independently of other factors such as discrimination to promotion, which leads to overestimating the explanatory part of human capital (Elder et al., 2010). This underestimation of women's human capital is further confirmed by the absence of any negative impact on women's employment when equal pay policies were applied, showing that they were paid below their marginal productivity under this theory.

Thus, to consider that controlling for human capital covariates offers us a "net" GWG, and that the difference between gross and net is "legitimate" and/or "fortuitous" and/or the result of a rational choice, is extremely questionable and renders invisible some of the GWG's causal mechanisms.

9.3 On the dimensions considered

Finally, it should be remembered that, while it is necessary to take into account various variables relating to women's situation on the labour market, i.e. not only wage differentials but also the Female Work Share, in order to identify the potential perverse effects of an equal pay policy on this share, this share must be qualified as a public policy objective. Indeed, the British policies of the 2000s, which encouraged mothers to return to the labour market quickly, both by increasing their earned income and by drastically reducing social transfers for them, increased participation rates at the potential expense of mothers' wellbeing (Bryson, 2005), particularly for single mothers, who were accused of creating a "reserve proletarian army" (Grover, 2005). So, increasing the Female Work Share is not necessarily normatively desirable in itself, and the social consequences of such moves need to be studied.

9.4 On wage distribution

Finally, this work focuses only on median GWG values (mean values having been used as a robustness check). However, the distribution of wages is also important (Harkness, 1996; Fortin

et al., 2017). This is an important issue, both in terms of inequality and its impact, and in terms of public policy design. Indeed, distortions at the bottom of the distribution can be considered more important in terms of social optimality than those at the top. And the impact of public policy can be underestimated if the distribution is not observed. The introduction in the UK of a statutory national minimum wage in 1999 led to some narrowing of the GWG at the bottom of the wage distribution (Robinson, 2002), but had little impact on the GWG further up the distribution (Dickens and Manning, 2004). It disproportionately benefited women who were more likely to be low paid, especially part-timers (Dex et al., 2000). Similarly, the impact of a policy can vary according to level of education. The impact of the 2003 Illinois Act on the lower end of the distribution and on part-timers is potentially significant. If so and where applicable, if it had no impact on demand for women’s labour, minimum wages can be a simple, transparent, easily enforceable and a powerful tool for redressing GWGs at the lower end of the wage distribution.

10 Conclusion

The aim of this study was to assess the impact of the 2003 Illinois Pay Transparency Act on discrimination against women in the private sector labor market, and more specifically on the Gender Wage Gap and Women’s Share Work, whose reduction could be a perverse effect of the reduction in the first variable of interest. However, the Pay Transparency Act was part of a broader Equal Pay Act. In the absence of any other state treated solely by an Equal Pay Act over the period 2000-2006, we were unable to create a design to isolate the effect of the Pay Transparency measure. We therefore estimated the total effect of the 2003 Equal Pay Act, including Pay Transparency, over the years 2003, 2004 and 2005 (there was no anticipation effect) using the Difference-in-Difference with fixed effects method. Given the fragility and unrealistic nature of the Common Trends hypothesis between Illinois and the other American states, even when reduced to a control group, we supplemented this with the Synthetic Control Method, by constructing a synthetic counterfactual Illinois, then carrying out placebo tests for the other control states. We tested median and mean values, and tested for the inclusion of public sector employees as well. Although the estimates indicate a non-trivial reduction effect of the GWG with no negative impact on the women’s work share, none of these estimates is statistically significant at the usual thresholds, for both methods. We cannot therefore infer an effect of this Equal Pay and Pay Transparency Act. But we must ask ourselves why, and there are several possible reasons: ignorance of the law, too little content and no real control mechanism, the psychological burden of initiating proceedings for the woman employee... And, while avoiding the file drawer effect, we must qualify the role of statistical significance, and bear in mind that the effect of a policy is highly complex. Beyond the external validity for Illinois, we need to conduct qualitative and psycho-sociological research on the impact of such measures on the labor market, Illinois included. Finally, it is not impossible that there has been a significant impact within the wage distribution, particularly among the lowest paid. There is much more research to be done, both in terms of evidence-based policy-making and in terms of a critical epistemological and methodological look at the tools considered, particularly the reification of human capital.

References

- Abadie, A. (2021). "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects." *Journal of Economic Literature*, 59 (2): 391-425.
- Altman, M. (2004). "Statistical significance, path dependency, and the culture of journal publication". *The Journal of Socio-Economics*, 33(5):651 – 663.
- Artz, B., Goodall, A. H., & Oswald, A. J. (2018). "Do Women Ask Industrial Relation"s, 57(4), 611–636.
- Andersson, J-S., and Ekman, G. (2018). "The impact of the Equal Pay Act of 2003 on gender equality in the Illinois labour market.", Bachelor's thesis in Economics, University of Gothenburg.
- Arrow, K. J. (1973), "The Theory of Discrimination", in O. Ashenfelter and A. Rees (eds.), *Discrimination in Labor Markets*, Princeton, NJ: Princeton University Press.
- Athey, S. and Imbens, G. (2022). "Design-based Analysis in Difference-In-Differences Settings with Staggered Adoption," *Journal of Econometrics*, 226 (1), 62–79.
- Baker, A., Larcker, D.F., and Wang, C.C.Y. (2022). "How much should we trust staggered difference-in-differences estimates?," *Journal of Financial Economics*, 144 (2), 370–395.
- Barth, E., Bryson, A., Davis, J., & Freeman, R. B. (2016). It's Where You Work: Increases in the Dispersion of Earnings across Establishments and Individuals in the United States. *Journal of Labor Economics*, 34(S2), S67–S97.
- Becker, GS. (1962). "Investment in human capital: A theoretical analysis." *Journal of Political Economy* ;70(5):9–49.
- Becker, GS. (1985). "Human Capital, Effort, and the Sexual Division of Labor", *Journal of Labor Economics*, Jan., 1985, Vol. 3, No. 1, pp. S33-S58.
- Bennedsen, M., Simintzi, E., Tsoutsoura, M., and Wolfenzon, D. (2018), 'Do Firms Respond to Gender Pay Gap Transparency?', NBER Working Paper No. 25435.
- Bertrand, M., Duflo, E., and Mullainathan, S. (2004). "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics*, 119(1), 249–275.
- Blau, F.D., and Kahn, L.M. (2017). "The gender wage gap: Extent, trends, and explanations". *Journal of Economic Literature*. 55(3):789–865.
- Blinder, AS. (1973). "Wage discrimination: Reduced form and structural estimates." *Journal of Human Resources* ; 8(4):436–455.
- Borusyak, K. and Jaravel, X. (2018). "Revisiting Event Study Designs," SSRN Scholarly Paper ID 2826228, Social Science Research Network, Rochester, NY.
- Brochier, C. (2020). "Mesurer la discrimination salariale ? Une critique radicale de la décomposition de Oaxaca-Blinder." *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 145(1), 61–79.
- Bryan, M. L., & Bryson, A. (2016). "Has performance pay increased wage inequality in Britain?" *Labour Economics*, 41, 149–161.
- Bryson, A. (2005), 'Working off Welfare', ch. 4 in H. Bochel, C. Bochel, R. Page, and R. Sykes (eds), *Social Policy: Issues and Developments*, Harlow, Pearson Education, 66–86.
- Bryson, A., Joshi, H., Wielgoszewska, B., & Wilkinson, D. (2020). A short history of the gender wage gap in Britain. *Oxford Review of Economic Policy*, 36(4), 836–854.

- Buchanan, J. (1996). "Minimum wage vs. supply and demand.", *Wall Street Journal*, April 24, 1996.
- Callaway, B. and Sant'Anna, P.H.C. (2021). "Difference-in Differences with multiple time periods," *Journal of Econometrics*, 225 (2), 200–230.
- Calonico, S., Cattaneo, M., and Titiunik, R. (2014). "Robust nonparametric confidence intervals for regression-discontinuity designs." *Econometrica*, vol. 82, no 6, p. 2295-2326.
- Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., and Wu, H. (2016). "Evaluating replicability of laboratory experiments in economics". *Science*, 351(6280):1433–1436.
- CAP (2021). "Women of Color and the Wage Gap", Center for American Progress.
<https://www.americanprogress.org/article/women-of-color-and-the-wage-gap/>
- Card, D. and Krueger, A. (1994). "Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania." *The American Economic Review* 84(4):772-793.
- Chamberlain, A. (2016). "Demystifying the Gender Pay Gap: Evidence From Glassdoor Salary Data." : <https://www.glassdoor.com/research/app/uploads/sites/2/2016/03/Glassdoor-Gender-PayGap-Study.pdf>
- Colella, A., Paetzold, R.L., Zardkoohi, A. and Wesson M.J. (2007). "Exposing Pay Secrecy." *Academy of Management Review* 32(1): 55– 71.
- CONSAD Research. (2009). "An Analysis of the Reasons for the Disparity in Wages Between Men and Women." Issued January 12.
- de Chaisemartin, C. and D'Haultfoeulle, X. (2020). "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 110 (9), 2964–2996.
- de Chaisemartin, C. and D'Haultfoeulle, X. (2022). "Two-way Fixed Effects and Differences-in-Differences Estimators with Several Treatments", NBER Working Paper.
- Deakin, S., Butlin, S. F., McLaughlin, C., and Polanska, A. (2015), 'Are Litigation and Collective Bargaining Complements or Substitutes for Achieving Gender Equality? A Study of the British Equal Pay Act', *Cambridge Journal of Economics*, 39, 381–403.
- Dex, S., Sutherland, H., and Joshi, H. (2000), 'Effects of Minimum Wages on the Gender Pay Gap', *National Institute Economic Review*, 173, 80–8.
- Dias, F. A., Chance, J. E., & Buchanan, A. (2020). The motherhood penalty and The fatherhood premium in employment during covid-19: evidence from The united states. *Research in Social Stratification and Mobility*, 69, 100542.
- Dickens, R., and Manning, A. (2004), 'Has the National Minimum Wage Reduced UK Wage Inequality?', *Journal of the Royal Statistical Society Series A*, 167(4), 613–26.
- Diprete, TA and Buchmann, C. (2006). "Gender-specific trends in the value of education and the emerging gender gap in college completion." *Demography* ;43(1):1–24.
- Duflo, E. (2001). "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review*, 91 (4): 795-813.
- Eisenberg, D.T. (2011). " Money, Sex and Sunshine: A Market-Based Approach to Pay Dis-

- crimination.” Mimeo. University of Maryland School of Law.
- Fortin, N., Bell, B., and Böhm, M. (2017), ‘Top Earnings Inequality and the Gender Pay Gap: Canada, Sweden, and the United Kingdom’, *Labour Economics*, 47, 107–23.
- Foster, T.B., Murray-Close, M., Christin Landivar, L., deWolf, M. (2020). ”An Evaluation of the Gender Wage Gap Using Linked Survey and Administrative Data”, CES WP 20-34, Women’s Bureau, U.S. Department of Labor.
- Fricke, H. (2017). ”Identification based on Difference-in-Differences Approaches with Multiple Treatments”, *Oxford Bulletin of Economics and Statistics*.
- Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in Competitive Environments: Gender Differences. *Quarterly Journal of Economics*, 118(3), 1049–1074.
- Goldin, C. (2014), ”A Grand Gender Convergence: Its Last Chapter”, *American Economic Review*, 104(4), 1091–119.
- Goodman-Bacon, A. (2019). ”Difference-in-differences with variation in treatment timing”, *Journal of Econometrics*, Volume 225, Issue 2, Pages 254-277.
- Gornick JC, and Jacobs JA. (1996). ”A cross-national analysis of the wages of part time workers: Evidence from the United States, the United Kingdom, Canada and Australia.” *Work Employment and Society*.10(1):1–27.
- Grembi, V., Nannicini, and T., Troiano, U. (2016). ”Do fiscal rules matter?.” *American Economic Journal: Applied Economics*, vol. 8, no 3, p. 1-30.
- Grover, C. (2005), ‘The National Childcare Strategy: The Social Regulation of Lone Mothers as a Gendered Reserve Army of Labour’, *Capital and Class*, 85, 63–90.
- Hamermesh, D.S. (1993), *Labor Demand*, Princeton, Princeton University Press.
- Harkness, S. (1996), ‘The Gender Earnings Gap: Evidence from the UK’, *Fiscal Studies*, 17(2), 1–36.
- ILO (2022). Pay transparency legislation: Implications for employers’ and workers’ organizations. International Labour Organisation, Geneva.
- Imai, K. and Kim, I.S. ”On the Use of Two-way Fixed Effects Regression Models for Causal Inference with Panel Data,” *Political Analysis*, 29 (3), 405–415.
- Joyce, R., Norris Keiler, A., and Ziliak, J.P. (2018), ‘Income Inequality and the Labour Market in Britain and the US’, *Journal of Public Economics*, 162, 48–62.
- Kim, M. (2015). ”Pay Secrecy and the Gender Wage Gap in the United States”. *Industrial Relations*, 54(4), 648–667.
- Kunze, A. (2008). ”Gender wage gap studies: Consistency and decomposition.” *Empirical Economics*;35(1):63–76.
- Kunze, A. (2017). The Gender Wage Gap in Developed Countries. Social Science Research Network.
- Imbens, G.M. and Wooldridge. J.M. (2007). “ What’s New in Econometrics? Lecture 10: Difference-in-Differences Estimation.”, NBER Working Paper.
- Lazear, E. P., and Rosen, S. (1990). Male-Female Wage Differentials in Job Ladders. *Journal of Labor Economics*, 8(1), S106–S123.
- Lips, HM. (2013). ”The gender pay gap: Challenging the rationalizations perceived equity, discrimination, and the limits of human capital models.” *Sex Roles* ;68(3):169–185.

- Ludwig, J., and Miller, D.L. (2007). "Does Head Start improve children's life chances? Evidence from a regression discontinuity design." *The Quarterly journal of economics*, vol. 122, no 1, p. 159-208.
- Manning, A., & Saidi, F. (2010). Understanding the Gender Pay Gap: What's Competition Got to Do with it? *Industrial and Labor Relations Review*, 63(4), 681–698.
- Mincer, J. (1974). *Schooling, experience and earnings*. Columbia University Press.
- Neuburger, J., Kuh, D., and Joshi, H. (2011), 'Cross-cohort Changes in Gender Pay Differences in Britain: Accounting for Selection into Employment Using Wage Imputation, 1972–2004', *Longitudinal and Life Course Studies*, 2(3), 260–85.
- Neumark, D., M. Schwarzer and W. Wascher. 2004. "Minimum Wage Effects Throughout the Wage Distribution", *Journal of Human Resources*, Vol 39, 425-453.
- Niederle, M., & Vesterlund, L. (2007). Do Women Shy Away From Competition? Do Men Compete Too Much? *Quarterly Journal of Economics*, 122(3), 1067–1101.
- Oaxaca, R. (1973). "Male-female wage differentials in Urban Labor markets." *International Economic Review* ; 14(3):693–709.
- Olsen, W., and Walby, S. (2004). "Modelling Gender Pay Gaps." Equal Opportunities Commission Working Paper Series No. 17. Manchester: Equal Opportunities Commission.
- Pekkarinen, T. (2012), "Gender differences in education." *Nordic Economic Policy Review*. 1:165–194.
- Polachek, S. (2019). "Equal pay legislation and the gender wage gap", *IZA World of Labor*, No 16v2, 16.
- Ponthieux S, and Meurs D. (2015). "Gender Inequality". In: Atkinson AB, Bourguignon F, editors. *Handbook of Income Distribution*. Elsevier; pp. 981–1146.
- Robinson, H. (2002), 'Wrong Side of the Track? The Impact of the Minimum Wage on Gender Pay Gaps in Britain', *Oxford Bulletin of Economics and Statistics*, 64(5), 417–48.
- Roth, J, Sant'Anna, H.C.P., Bilinski, A., Poe, J. (2023). "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature". On arxiv : <https://arxiv.org/pdf/2201.01194.pdf>
- Rotman, A., and Mandel, H. (2023). "Gender-Specific Wage Structure and the Gender Wage Gap in the U.S. Labor Market." *Social indicators research*, 165(2), 585–606.
- Ryan, AM., Burgess, JF. Jr., Dimick, JB. (2015). "Why We Should Not Be Indifferent to Specification Choices for Difference-in-Differences". *Health Serv Res*. 50(4):1211-35.
- Self, S.G., and Mauritsen, R.H. (1988). "Power/Sample Size Calculations for Generalized Linear Models", *Biometrics*, Vol. 44, No. 1, pp. 79-86 (8 pages).
- Snow, J. (1855). *On the Mode of Communication of Cholera*. London: John Churchill, New Burlington Street, England.
- Strezhnev, A. (2018). "Semiparametric weighting estimators for multi-period difference-indifferences designs," Working Paper, 2018.
- Sun, L. and Abraham, S. (2021), "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 225 (2), 175–199.
- Vaccaro, G. (2017), 'Using Econometrics to Reduce Gender Discrimination: Evidence from a Difference-in-discontinuity Design', mimeo.

- Webber, D. (2016b). Firm-Level Monopsony and the Gender Pay Gap. *Industrial Relations*, 55(2), 323–345.
- Weintraub, P.G. (2016). "The Importance of Publishing Negative Results", *Journal of Insect Science*, Volume 16, Issue 1, 109
- Zabalza, A., and Tzannatos, Z. (1985), 'The Effect of Britain's Anti-discriminatory Legislation on Relative Pay and Employment', *The Economic Journal*, 95, 679–99.
- Yasuda, H. (2023). Employers' stereotypes and taste-based discrimination. *Journal of International Economy*, 67, 101240.
- Yu, W., & Hara, Y. (2021). Motherhood Penalties and Fatherhood Premiums: Effects of Parenthood on Earnings Growth Within and Across Firms. *Demography*, 58(1), 247–272.
- Zeldow, B., and Hatfield, L.A. (2021). "Confounding and regression adjustment in difference-in-differences studies". *Health Serv Res*, 56(5):932-941.
- Ziliak, S. T. and McCloskey, D. N. (1996). "The standard error of regressions." *Journal of Economic Literature*, 34(1):97–114.
- Ziliak, S. T. and McCloskey, D. N. (2004). "Size matters: the standard error of regressions in the american economic review". *The Journal of Socio-Economics*, 33(5):527 – 546.
- Ziliak, S. T. and McCloskey, D. N. (2008). *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. University of Michigan Press, Ann Arbor, Mich.