Série des Documents de Travail

n° 2014-47

Does Endogenous Matching Explain the Family Pay Gap? Evidence from Linked Employer-Employee Data

L.WILNER[1]

[1] CREST (INSEE). E-mail: lionel.wilner@ensae.fr

Does Endogenous Matching Explain the Family Pay Gap? Evidence from Linked Employer-Employee Data*

Lionel Wilner†

February 2015

Abstract

Once individual unobserved heterogeneity and human capital depreciation have proved not to fully account for wage differences consecutive to parenthood, a remaining explanation (discrimination aside) could be that parents select into low wage firms. This paper tests that hypothesis by resorting to linked employer-employee data and by introducing firm fixed effects in hourly wage equations, on top of worker fixed effects and usual measures of human capital. A motherhood penalty remains despite the presence of two-way high dimensional fixed effects. By contrast, men do not experience any loss due to childbirth, but they do not enjoy any premium either.

Keywords: High dimensional fixed effects; worker fixed effects; firm fixed effects; linked employer-employee data; wages; gender.

JEL Classification: J12; J13; J16; J31; J62; J71.

---

*I thank Richard Blundell, Élise Coudin, Francis Kramarz, Thomas Le Barbanchon, Laurent Linnemer, Thierry Magnac, Johannes Spinnewijn, Sébastien Roux, Michael Visser and Andrea Weber for their insightful comments. All errors and opinions are mine.

†CREST (INSEE). 18, boulevard Adolphe Pinard, 75 014 Paris, France. Tel: (+33)141175460. Email: lionel.wilner@ensae.fr

1 Introduction

"Qui va garder les enfants maintenant ?"
"Who's gonna watch the kids now?"
Presidential campaign, France, 2007.

As both a woman and a mother, the candidate Ségolène Royal faced a lot of sexist remarks during the 2007 campaign for French presidential elections. Surprisingly, the toughest critiques came from her own party (the left-wing socialist party), as noticed by most political observers. The quotation above apparently refers to the fact that Royal's companion back then was François Hollande.[1] It shows not only how deeply grounded sexist stereotypes are, but also how reluctant some individuals (presumably men, but not only) still are to be ruled by women.[2]

Not only do gender inequalities remain within households (in terms of the share of domestic work or bargaining power, see e.g. Meurs and Ponthieux, 2014), but they also persist within firms, which both labor economists and sociologists have been documenting for decades. The gender pay gap, occupational gender segregation and the glass ceiling are most striking examples: they are both unfair and inefficient as long as they are not justified by productivity differentials. Even if the gender gap has begun to reduce slightly, it does not fall to zero (OECD, 2012). Composition effects tend to locate women at lower positions in the hierarchy, which contrasts with their higher level of education on average –for most members of the European Union at least. This discrepancy between education and occupation has been designated as a form of gender segregation. In the same vein, the glass ceiling that prevents women from reaching top positions of the institutional hierarchy (CEOs, States' presidency, etc.) is hard to break.

However the most obvious gender inequality is related to childbirth. The family pay gap, which accounts for hourly wage differentials consecutive to childbirth, has been first documented by Waldfogel (1997). Numerous papers then determined

[1]Royal is currently Minister of Ecology and Sustainable Development while Hollande is President of the French Republic.

[2]The author of that quotation ignores probably that more than two centuries went by since Mary Wollstonecraft condemned the restrictive domestic sphere to which women were confined. In Wollstonecraft (1792), she argued for the same level of education for men and women and, implicitly, for no other discrimination than capacities. Ironically, the posterity has especially remembered her private life as storied *post mortem* by her husband, the philosopher William Godwin, in Godwin (1798). Mary Wollstonecraft was (also) the mother of the writer Mary Shelley.

the magnitude of motherhood penalties. More recent papers have also assessed the existence of wage differences among men between fathers and non-fathers. Four theoretical explanations for a hourly wage gap have mainly been proposed, namely (i) human capital depreciation (ii) individual unobserved heterogeneity (parents would have lower average productivity) (iii) firm matching (parents would match low-wage firms) (iv) discrimination. Yet the relevance of each of these explanations remains to be assessed empirically. Moreover, and to the best of our knowledge, the firm matching explanation has never been investigated.

The contribution of this paper consists in testing whether the family pay gap stems from selection effects of parents into firms, or endogenous matching. For this purpose, we resort to panel data that are linked employer-employee data (LEED), that is, the statistical unit of which is a (individual, firm, year) triplet. We estimate an adjusted family pay gap thanks to a two-factor model with high dimensional worker and firm fixed effects, which permits to control for as much observed and unobserved differences as possible in both individual and firm characteristics. Our application is based on the exhaustive DADS panel which contains comprehensive information on French salaried employees' careers from 1995 to 2011 in the private sector. LEED have proved exceptionally useful in the estimation of wage equations since Abowd, Kramarz, and Margolis (1999). Up to now, this paper sounds like the first attempt to analyze the family pay gap by resorting to such data, and thus to a specification that allows for firm heterogeneity to be an explanatory factor for hourly wage differentials. By disentangling the effect of childbirth from spurious correlation between parenthood and other firm-specific wage determinants, this estimation contrasts with previous analyses that control for individual unobserved heterogeneity only.

After controlling for full-time and part-time experience as well as for both firm and worker fixed effects, we still find a difference between mothers and non-mothers which amounts to about -3% per child on the hourly wage, while there is no significant difference for men. However these results must not be interpreted as the sign of any discrepancy with the idea of a fatherhood premium: they are rather consistent with heterogeneity in those premia, which previous literature has been documenting. Since we do not find significant differences with the estimation that omits to control for firm fixed effects (especially in the case of women), we reject the "firm matching" explanation. This provides indirect evidence of either long-run, dynamic effects, consistently with recent findings by Kleven, Landais, and Søgaard (2015), or of gender discrimination against mothers at work. Finally, we propose an evaluation of the contribution of the family pay gap to the gender gap, by simulating a counterfactual *scenario* in which women and men experience the same childbirth penalty.

The rest of the paper is organized as follows. Section 2 is devoted to a literature review. Section 3 presents our matched employer-employee database. We expose our econometric specification in Section 4. Section 5 contains our results, namely the test of the firm matching explanation, as well as a measure of the contribution of the family gap to the gender gap, and robustness checks. Section 6 concludes with some policy recommendations.

2 Literature review

The seminal contributions of Waldfogel (1997, 1998) document the existence of a motherhood wage penalty both in the US and in the UK. Relying on data from 1968 to 1988, and resorting to Mincer equations on log hourly wages with individual fixed effects, she finds a wage loss of about -6% per child. On a different period of estimation, from 1982 to 1993, Budig and England (2001) estimate this loss at -3% for the first child, -9% for the second child and -12% for the third child. Similar figures are found in other European countries: an exhaustive list is given in Davies and Pierre (2005). In Germany, the motherhood wage penalty turns out to be higher than 10% in absolute: Buligescu, De Crombrugghe, Menteşoğlu, and Montizaan (2009) find a family pay gap between -10% and -14%; Beblo, Bender, and Wolf (2009) report estimates higher than 19% in absolute; according to Felfe (2012), this gap approximates -10.7%. More recently, Kleven, Landais, and Søgaard (2015) study the case of Denmark and find long-run, dynamic effects of childbirth on mothers' wages.

The French case has been investigated by Meurs, Pailhé, and Ponthieux (2010). They focus on the effect of spending child-related time out of the labor market on the gender pay gap (previously documented in Meurs and Ponthieux, 2000). Lequien (2012) and Joseph, Pailhé, Recotillet, and Solaz (2013) analyze more specifically the impact of parental leave by exploiting two reforms that enable them to recover causal effects of time out of the labor workforce on wages. The first reform occurred in 1994 and is related to a monthly benefit for parents called the *Allocation Parentale d'Education* (APE); it changed the incentives to take a parental leave after the birth of a second child by extending allowance eligibility from third born to second born child. The second reform is the creation of a childhood benefit that replaced the APE, the *Prestation d'Accueil du Jeune Enfant* (PAJE); the authors exploit a specific feature of this reform: the attribution of some supplemental benefit for part-time activity.

Some papers nevertheless find no wage differential between mothers and non-mothers. On US data, Korenman and Neumark (1992) do not find any evidence of a family pay gap. However, their analysis relies on first-differences over a short period, from 1980 to 1982, which could explain this dissonance. More convincing is the paper by Simonsen and Skipper (2006) that finds a "net gap" (unadjusted) but no "direct gap", *i.e.*, causal effect, of childbirth on wages. The authors explain that most of the gap could stem from indirect channels relating motherhood to other covariates, which may cause spurious correlation. For instance, motherhood may be negatively correlated with experience since child-rearing activities lead mechanically to lower work experience. In the same vein, one expects to find more mothers working in the public sector or in more family-friendly occupations. When looking for causal effects, these aspects have to be controlled for. Using a propensity-score matching approach, the authors conclude that in Scandinavia, mothers self-select into the public sector (see also Nielsen, Simonsen, and Verner, 2004) and that there is no causal effect of motherhood on wages in the private sector.

To recover the causal effect of childbirth on wages, it has been thought of using data on twin sisters, which provide with a natural experiment. Comparing wage trajectories of mothers with respect to non-mothers' ones, Lundberg and Rose (2000) find a causal wage gap of -5%, a figure close to the one obtained by Simonsen and Skipper (2012) following a similar approach. Another method consists in relying on exogenous childbirths, by exploiting for instance fertility shocks. The introduction of contraceptive methods and the passing of abortion laws provide researchers with quasi-experimental settings that give the variation required for identification. Miller (2011) uses biological fertility shocks as instruments for age at first birth to recover a causal impact of the timing of childbirths on wages: delaying birth by one year would raise earnings by up to 9/10% and increase work hours by 6%.

A recent strand of research has been focusing on the heterogeneity of the motherhood wage penalty, according to either the rank in the wage distribution or the level of education. A controversy is currently opposing Budig and Hodges (2010) to Wilde, Batchelder, and Ellwood (2010) about the link between education and this wage differential: the former find higher motherhood penalties for women with higher cognitive skill levels, while the latter obtain higher penalties at lower wage levels. England, Bearak, Budig, and Hodges (2013) address this issue and try to reconcile both points of view by introducing other dimensions of heterogeneity like race in the US.

Another contemporaneous area of research has begun to focus on men and has investigated the issue of a fatherhood wage gap. Contributions include Lundberg and Rose (2000, 2002) and Glauber (2008). Results conclude rather to the existence of a fatherhood premium, which contrasts with the motherhood penalty and constitutes a gender-based inequality with respect to childbirth. However, Killewald (2013) finds much heterogeneity in those premia: certain groups of fathers, including unmarried residential fathers, nonresidential fathers and stepfathers experience no significant premium at all.

The family pay gap is a puzzling issue and several theoretical explanations have been proposed to account for that wage differential.

First, motherhood implies some human capital depreciation due to the mandatory parental leave. This "human capital deterioration" explanation dates back to Becker (1985). Human capital is a composite concept: it aggregates at least education, experience and training. Women who want to become mothers would rationally opt for lower educational levels. Their career has mechanically more frequent interruptions (sick leaves, either for themselves or for their children, on top of the maternity leave), which depreciates their work experience. Furthermore, the time spent out of the labor workforce is likely to have a negative impact on their training, especially if training is some function of continuous employment. Last but not least, mothers may prefer to work part-time, which further diminishes their work experience.[3] Under the assumption of perfectly competitive labor

[3]One could argue that working part-time constitutes a bad signal sent to their employers by individuals reducing voluntarily their activity. However, this explanation does not belong to the "human capital" theory, but rather to a competing explanation, the "signalling" theory proposed by Spence (1973).

markets, lower hourly wages must reflect a lower productivity caused by career interruptions, a lower training or a lower education level. In Sweden, Albrecht, Edin, Sundström, and Vroman (1999) ask whether this hypothesis could alone be responsible for the family pay gap; they find that human capital depreciation is not the sole explanation for the negative effect of career interruptions on subsequent wages.

Second, individual unobserved heterogeneity has been invoked to explain the wage gap between mothers and non-mothers. The formers may choose more family-oriented careers, this self-selection being primarily driven by preferences and/or personal abilities. The negative correlation between labor market outcomes and fertility could then reflect either stronger preferences for family, domestic activities, leisure, or a lower on-the-job productivity. Women endowed with such preferences and capacities would *ex ante* invest less in education and training, hence acquiring less human capital. The family gap could thus reflect a different willingness to work in a competitive environment. The capacity to disentangle spurious correlation between childbirths and wages due to preferences from the causal effect of motherhood requires therefore to control for unobserved heterogeneity, for instance thanks to individual fixed effects –assuming that unobserved heterogeneity does not vary over time. Even after controlling for worker fixed effects on panel data, as is the case in most empirical papers cited previously, a substantial part of wage differences remains between mothers and non-mothers.

A third explanation (hereinafter "firm matching") claims that mothers (or more generally parents) would match less productive firms. To reconcile both family life and career, women who ought to be mothers would look for jobs that allow them to spend more time at child-rearing activities. For instance, they would favor jobs with flexible hours, on-site day care, jobs in which personal phone calls are authorized during work, or jobs that do not require overtime work, evening work, work during weekends, etc. In that case, occupational segregation should emerge at the equilibrium on the labor market. As a result, forward-looking women who want children would seek more convenient and less energy-intensive jobs. In the same vein, mothers may have higher search costs on the labor market, which restricts their mobility and prevents them from looking for better positions while resulting in poor job matches. Surprisingly this explanation has received little attention so far. Budig and England (2001) proposed to control for as many job characteristics as possible, including part-time, in order to neutralize any "family-friendly" job feature. From the huge literature devoted to job search we know that there is a strong heterogeneity in the quality of the employer-employee relationship, and that mobility offers potentially large wage gains. Nielsen, Simonsen, and Verner (2004) show that mothers tend to self-select into the public sector. Beblo, Bender, and Wolf (2009) argue that the selection into private establishments has to be taken into account since it could represent up to 7pp of the family pay gap in the German case (-19% after controlling for establishment effects in a matching approach against -26% when such effects are ignored). This explanation has also been put forward by Felfe (2012) who suggests that mothers are ready to trade off earnings against amenities through a compensating wage differentials (CWD) story. Among mothers she distinguishes those who maintain themselves into the same position after maternity leave from the others. She finds that the former experience a significantly smaller wage gap (-9.3% against -24.3%) than the latter, which support this hypothesis. However, part of the difference stems precisely from an adjustment of work conditions, and after controlling for this adjustment the family pay gap cannot be solely explained by a CWD story.

Fourth and finally, a remaining explanation for the motherhood wage penalty is the possibility of discrimination against mothers at work. Employers could be reluctant to hire mothers-to-be or women that they expect to become mothers, which would affect mothers' employment. On top of that, they could also be more severe in the wage bargaining process and offer prospective mothers less chances of distinguishing themselves (through the provision of overtime work, more risky missions, etc.), which would result in a less frequent attribution of irregular bonuses. Generally speaking, such a discrimination could result from job reallocation, either within firm or within establishment. The strategy adopted in this paper is to bring indirect evidence in favor of discrimination by testing the three previous explanations and ruling them out. The elimination of alternative explanations, combined with the absence of more convincing hypotheses so far, suggests that gender-biased discrimination is likely.

3 Data

3.1 Sources

Our analysis is based on the merger of two French administrative datasets commonly known as the DADS-EDP panel collected by Insee.

The first source is the DADS panel, a comprehensive database of salaried employees, and the longitudinal version of the cross-sectional *Déclaration Annuelle de Données Sociales* (DADS). This panel is exhaustive: it is mandatory for French firms to fill in the DADS annually for every employee subject to payroll taxes, which encompasses all salaried employees in the private sector or in government-owned firms (the so-called "semi-public sector"), at the exclusion of civil servants and independent workers. From 1967 to 1975, the panel does not contain any information on firms while from 1976 to 2011, the data are available at the individual-firm level. Every firm – more precisely every establishment – has a unique identifier, the SIRET,[4] a 14-digit number, while individuals are identified by their NIR, a social security number with 13 digits. Before 1976 an observation is made up of a unique (individual, year) pair while after 1976 it is composed of a unique (individual, firm, year) triplet, which features the data as linked employer-employee data (LEED). From another source specific to the public sector that was merged into the DADS panel, we dispose also of information on the public sector, more precisely on its three distinct components: a nation-wide part (State civil servants), a local component,[5] and public hospitals' employees. The data about State civil servants have been available since 1978, while the local component was introduced in 1988; public hospitals have been included since 1984. The DADS panel contains information about individuals born on October of even-numbered years –a representative sample of the French salaried population at rate 1/24. From 2002 onwards though, the panel has been completed with individuals born on October of odd-numbered years, which gives a sampling rate of 1/12; however, longitudinal depth is mechanically shorter for such individuals in comparison with those born on October of even-numbered years. Since filling in the DADS form is mandatory,[6]

[4]The SIRET is a concatenation of the SIREN, a firm identifier, and of an establishment identifier.

[5]France is divided into several administrative layers (regions, *départements* or counties, municipalities, etc.) which all employ their own civil servants.

[6]The absence of a DADS as well as incorrect or missing answers are punished by law with fines.

and because of its comprehensiveness, the data are of exceptional quality and have low measurement error in comparison with survey data. Some years are missing: 1981, 1983 and 1990 because there was no data collection by Insee during the 1982 and 1990 censuses. In 1994, 2003, 2004 and 2005, the quality is nevertheless questionable. Since overseas appeared in the panel from 2002 onwards, we restrict our attention to metropolitan France. Finally, these data contain detailed information about gross and net wages, work days, work hours, other job characteristics (from 1976 onwards: the beginning and the end of an employment's spell, seniority, a dummy for part-time employment), firm characteristics (industry, size, region) and individual characteristics (gender, age).

The second source is the *Échantillon Démographique Permanent* (hereinafter EDP). This longitudinal database covers a representative part of the population born on one of the first four days of October. It contains administrative registers of births and marriages from 1968 to 2011, as well as partial information on education[7] from 1968, 1975, 1982, 1990 and 1999 censuses. However, for half of the sample, namely people born on October 2^{nd} or 3^{rd}, birth registers have not been properly filled in from 1983 to 1997. The information is incomplete from 1983 to 1989 and missing from 1990 to 1997. It is possible[8] to recover part of missing data by exploiting 1990 and 1999 censuses, but at the cost of introducing some measurement error since the information conveyed by both censuses does not coincide perfectly. We choose the most conservative option: we rely on birth registers and keep therefore individuals born on October 1^{st} or 4^{th} only. This extract from the EDP is still representative of the French population[9] at rate 2/365.

The two sources can be merged thanks to their common individual identifier, the NIR. We exclude "wrong" NIRs[10] which are only present in the DADS panel for cross-sectional use and for statistical reasons, as we cannot follow careers of such individuals. We also exclude self-employed persons who appeared in the panel from 2009 onwards. Finally, we eliminate observations that correspond to home-work or to unemployment amenities.

A methodological contribution of this paper consists in computing an accurate measure of salaried experience. We exploit the administrative feature of our

[7]See Footnote 18.
[8]An approach probably followed by Lequien (2012).
[9]Individuals born abroad are also missing from the EDP.
[10]For instance, some of them were not born on October.

dataset and derive experience at the individual level from the sequence of observed working times. We count the actual (past and present) number of working hours and convert it in full-time (full-year, or also full-time equivalent) units. In France, a full-time worker used to work 2028 (1820) hours per year before (after) 2002 –the mandatory working time decreased by 4 hours per week after the adoption of Aubry laws. However, we face data limitations: worked hours have been available since 1995 only. Hence we pay attention to individuals who entered the panel after 1995 in order not to bias our measure of experience because of missing or incomplete sequences of working time. After merging the two datasets and imposing this "entry condition", our sample includes 46,280 individuals. We proceed then to further and more innocuous selection described more extensively in Appendix A: we focus on individuals aged 16 to 65 working in the private sector and whose annual wage exceeds 10 euros in 2011 terms. We disregard years 2003 to 2005 in our analysis for the reasons mentioned above. Our working sample contains 41,531 individuals, which represents 212,189 observations at the individual-year level and 301,079 observations at the individual-firm-year level. Even if our sample is not representative of the whole French salaried population, the method we propose to deal with the family gap issue applies to any LEED sample.[11]

3.2 Descriptive analysis

Table 1 provides a summary description of our working sample. Our individuals are mechanically younger: they are aged 26.9 on average. 48% of them are women, 18.9% are married, 33.9% have at least one child. Some individuals have been continuously working in the private sector from 1995 to 2011. The mean potential experience amounts to 6.4 years, where potential experience is defined as the difference between the current year and the year of first appearance in the panel. On average the full-time experience is 2.3 years – in full-time full-year units – while the part-time experience amounts to 0.8 year. Average seniority is about 2.5 years, where seniority is the difference between the current year and the year of first appearance in the current firm. The annual job duration amounts to 237 days, or 1037 hours, which reflects composition effects explained by both youth insertion in the labor market and part-time activity. The average net annual wage

[11] Furthermore, as time goes by the entry condition becomes less restrictive –hence the selection will be less drastic in future work relying on the same source.

is 11050 euros in 2011 terms while the net hourly wage amounts to slightly less than 10 euros. Since we proceed to some trimming of very low hourly wages (see Appendix A and Section 5.3), the minimum observed hourly wage is 3.51 euros.

The issue of part-time work cannot be disregarded in such an empirical analysis devoted to childbirth as argued for instance by Budig and England (2001). By contrast to studies relying on full-time workers only, part-time workers are not selected out of our sample. However there is a remaining interrogation in the literature about whether part-time induces a penalty or a premium on hourly wage. On Australian data Booth and Wood (2008) find that the negative coefficient of part-time in a Mincerian equation on the hourly wage disappears once covariates (especially experience) and unobserved heterogeneity have been controlled for. We adopt temporarily their methodology and reproduce their Table 2.[12] Hence we specify:

$$\log W_{it} = X_{it}'\beta + \alpha P_{it} + \theta_i + \epsilon_{it}, \tag{1}$$

where W_{it} is the log hourly wage, X_{it} a set of covariates, P_{it} a dummy for part-time work, θ_i an individual fixed effect and ϵ_{it} an error term. The part-time dummy is defined as the fact of working less than 90% of the full-time equivalent –we check for the robustness of our results to the latter threshold. Table 2 reports the estimates of α under different specifications (with or without fixed effects θ_i, with or without covariates including worker characteristics, firm characteristics, experience, etc.). In both pooled OLS and fixed effects approaches, the sign of α becomes positive after controlling for observed and unobserved heterogeneity. This empirical result holds both for women and men.[13]

Another empirical issue worth examining is the relationship between childbirths and experience.[14] Once again it is crucial to distinguish among full-time and part-time experience. We thus specify:

$$\text{Experience}_{it} = X_{it}'\beta + \theta_i + \epsilon_{it}, \tag{2}$$

where Experience may refer either to full-time experience (measured in full-time full-year units) or to part-time experience. We do not include part-time work P_{it} as a covariate here for it would be correlated with determinants of the dependent

[12] I thank Sébastien Roux for a suggestion of this kind.
[13] In what follows we will not distinguish P_{it} from other covariates in X_{it}.
[14] I am most grateful to Richard Blundell for this suggestion.

variable, hence endogenous in (2). Table 3 displays the results which exhibit interestingly gender differences. While women may experience a loss of up to 3 years of full-time experience following late childbirths (after the fourth), men's experience never decreases when becoming fathers: on the contrary, they keep gaining years of full-time experience. Labor supply decisions within households consecutive to parenthood may still be strongly biased in favor of men's pursuing their activity on the labor market and women's reducing theirs. The diagnosis is altered as regards part-time experience since women's part-time experience increases after childbirth, which reflects some trade-off of full-time against part-time work. Yet the loss in full-time experience is far from being compensated by the raise in part-time experience –especially after the third childbirth. By contrast, men do not choose significantly more to work part-time upon childbirth –except after the third child. These results confirm the hypothesis of gender-biases in labor supply decisions within households, and suggest that staying out of the labor workforce is still frequent for women after the third childbirth. Finally, marriage is rather associated with positive impacts on full-time experience – higher for men than for women – but has no significant effect on part-time experience.

4 Econometric specification

4.1 Worker fixed effects

The literature devoted to the family pay gap has focused so far on the estimation of Mincerian wage equations on panel data at the individual-year level. The dependent variable usually considered is the logarithm of the hourly wage, viewed as a proxy for productivity if labor markets are perfectly competitive. Explanatory variables include experience, seniority, job characteristics (sector, firm size, location), possibly other controls, as well as time and individual fixed effects. In that vein we estimate first a linear model and specify for individual i observed on year t:

$$\log W_{it} = X_{it}'\beta + N_{jt}'\delta + V_t + \theta_i + \epsilon_{it}, \tag{3}$$

where W_{it} is the ratio between the sum of wages perceived by i in year t and the sum of his worked hours, X contains part-time activity, marriage and children as well as quadratic specifications in full-time experience, part-time experience and seniority,[15] N includes firm characteristics (size, industry, *département*) and ϵ_{it} is an idiosyncratic error term, the variance of which is allowed to be individual-specific. Our variables of interest are Childbirth_{itk}, $\forall k = 1, \ldots, 5$, which are dummy variables indicating whether individual i had already experienced his k-th childbirth at time t. We introduce the covariate Marriage_{it} to capture the effect of marriage on hourly wages. Moreover, we control for many job characteristics which correspond to the main employment's characteristics (see Appendix A for the definition of main employment) including the firm's size, the *département* where the establishment is located and its sector of activity. Size is coded with 12 categories while the industry is defined by the first two-digit of the NACE classification and has 39 categories, including a "missing" one. 95 *département* dummies account for the spatial dispersion of earnings in metropolitan France. V_t is a year fixed effect that captures the conjuncture affecting earnings (business cycle, macro shocks, etc.) while θ_i is an individual fixed effect that encompasses permanent unobserved heterogeneity including talent, employability, cohort effects but also initial education. The EDP provides us with a schooling variable that indicates the highest degree obtained by an individual. However, in the presence of individual fixed effects the coefficient of education is identified provided that this variable is time-varying, which is not the case with initial formation. Finally, in the spirit of what is done in Waldfogel (1997, 1998) and Budig and England (2001), we compare a fixed effect estimation with a first-difference estimation of model (3). While the former enables us to recover rather long-run effects, the latter accounts for a short-run effect.

4.2 Worker and firm fixed effects

Despite their qualities, previous models suffer from an important omitted variable bias caused by the omission of firm fixed effects, which has first been emphasized by Abowd, Kramarz, and Margolis (1999). High wage workers are likely to match high wage firms; more generally the matching process may allocate specific workers into firms with specific compensation schemes. If the job matching results in important selection effects, previous estimations suffer from an endogeneity bias due to a correlation between explanatory variables like Marriage_{it} or Childbirth_{itk}, $\forall k = 1, \ldots, 5$, and a firm-specific term ψ_j of the error term in (3) that would write

[15] These characteristics are attached to an individual's main employment.

in fact $\psi_j + \epsilon_{ijt}$. Abowd, Kramarz, and Woodcock (2008) explicit this omitted variable bias as a function of the covariance between the matrix X and the design matrix of indicator variables for the employer for which individuals work, conditional on the design matrix of individuals' indicator variables. We choose then to exploit the LEED nature of our dataset without aggregating information at the individual-year level. To the best of our knowledge, this paper is the first attempt to address the family pay gap issue at the individual-firm-year level and hence avoids the omitted variable bias due to endogenous matching. We specify a model *à la* Abowd, Kramarz, and Margolis (1999) in which we are able to identify firm fixed effects:

$$\log w_{ijt} = x'_{ijt}\beta + v_t + \theta_i + \psi_{j(i,t)} + \epsilon_{ijt}, \tag{4}$$

where w_{ijt} is the hourly wage earned by individual i working in firm j on year t, x contains part-time activity, marriage and children, as well as quadratic specifications in full-time experience, part-time experience and seniority, v_t is a year fixed effect, θ_i is an individual fixed effect, $\psi_{j(i,t)}$ is a firm fixed effect and ϵ_{ijt} an error term, the variance of which is individual-specific. We do no longer control for location, size or industry since these covariates correspond to an aggregation of the pure firm effects $\psi_{j(i,t)}$ (Abowd, Kramarz, and Woodcock, 2008); more precisely, they are an employment-duration weighted average of the firm effects within the *département*/size[16]/industry.

4.3 Identification

The identification of the model is discussed in Abowd, Kramarz, and Margolis (1999) and provided in longer details in Abowd, Creecy, and Kramarz (2002). It proceeds from connectedness properties of the graph formed by individuals (let us designate their number by N) and firms (let us designate their number by J). More specifically, the data must be partitioned into G mutually exclusive groups of either individuals or firms such that the members of one group cannot have employed – or have been employed by – any member of another group. These G groups are the maximally connected sub-graphs of the entire graph, the vertices of which correspond to the union of the set of persons and the set of firms, while its edges are pairs of firms and persons. For each group g with N_g persons and J_g

[16]By abuse of language, size refers to all firms with a size belonging to one of the 12 previously mentioned size categories.

firms, $N_g - 1$ individual fixed effects and $J_g - 1$ firm fixed effects can be identified so that $N + J - G$ effects can be identified on the whole. The uniqueness of the effects within a group stems from the elimination of one person effect: it can be achieved by setting the group mean to zero as Abowd, Creecy, and Kramarz (2002) suggest.

4.4 Estimation

Technical details regarding the estimation of two-way high dimensional fixed effects are provided in Abowd, Creecy, and Kramarz (2002); in particular, one practical solution to deal with the inversion of large matrices consists in exploiting their sparsity. Efficient algorithms include the conjugate gradient and the "zigzag" Gauss-Seidel routine.

Among the four explanations for the parenthood pay gap presented in Section 2, this two-factor model[17] enables us to distinguish carefully among individual unobserved heterogeneity and firm matching. First, to document selection effects, we recover the estimated individual fixed effect $\hat{\theta}_i$ and explain it thanks to cohort effects, education, marriage and parenthood as follows:

$$\hat{\theta}_i = \gamma_0 + \gamma_1 \text{Married}_i + \gamma_2 \text{Children}_i + \gamma_z Z_i + \eta_i, \tag{5}$$

where Married_i is a dummy which equals 1 if individual i has ever been married, Children_i are dummies which equal 1 if individual i has ever had children and Z_i includes education dummies[18] as well as cohort dummies. Put in other words, we seek here to explain unobserved heterogeneity by observed heterogeneity and project the estimated fixed effects on covariates. γ_1 measures how much married individuals differ from non-married individuals in terms of average hourly wage, while γ_2 measures how much parents differ from non-parents in this respect. Importantly there must not be any confusion between γ_1 and β_m on the one hand, and between γ_2 and $\beta_{ck}, \forall k = 1, \ldots, 5$ on the other hand. The γ coefficients describe how much married individuals (parents) differ with respect to singles (non parents) in terms on θ_i, that is, in terms of permanent hourly wage (or productivity). On the contrary, the β coefficients depict the effect of a time-varying event

[17] The two factors correspond to worker and firm fixed effects.

[18] Including a "missing" category when no information is available about the highest degree obtained, as is the case for 18277 individuals, namely 44% of the sample.

like marriage (childbirth) on wages.

Second, to assess the existence of an endogenous matching process that would match individuals into specific firms, we compute the correlation between individual and firm fixed effects $cor(\theta, \psi)$ that indicates the extent to which high wage workers self-select into high wage firms. Separate correlations for parents and for non-parents shed some light on the differences between the assignment of parents into high wage firms from the assignment of non-parents into high wage firms. For instance, it is well-known that such a correlation is almost zero in the US while it is negative in France –our findings are consistent with the latter result.

In practice, we estimate different models for women and for men, which allows us to proceed to separate analyses on both genders with respect to the issue of parenthood penalty. Once again, relatively little attention has been paid to men, and *a fortiori* to both genders at the same time, which is another dimension this paper contributes to.

5 Results

5.1 Testing for endogenous matching

Our main results are displayed in Table 4, columns 3 and 6, which report the estimates from the model including individual- and firm- fixed effects (2FE) with $G = 4,742$ groups for women and $G = 5,268$ groups for men. For both genders we estimate three different specifications: first-difference (FD) in columns 1 and 4, individual fixed effects (FE) in columns 2 and 5 as well as individual- and firm-fixed effects in columns 3 and 6.

Overall, and in line with previous findings, our estimations suggest the existence of a parenthood wage penalty for French women working in the private sector, about -3% per child. However, nothing significant is obtained as far as French men are concerned: no fatherhood premium is observed. The motherhood penalty exhibits some non-linearity with the rank of birth: -4.2% for the first child, -6.5% for the second child, -7.1% for the third child and -9.7% for the fourth child. Estimates for the fifth childbirth are much more imprecise due to low sample size. These results claim for the existence of a gender-bias in the relationship between children and wages. They are consistent with heterogeneity in childbirth returns (our sample is made up of young individuals), but also with long-run or dynamic effects of motherhood. The FD approach tends to estimate rather a short-run effect; interestingly the short-run motherhood penalty turns out to be systematically lower in absolute than the long-run loss measured thanks to FE or to 2FE methods, which is consistent with cumulative penalties and dynamic, long-run effects of childbirth found in Kleven, Landais, and Søgaard (2015). In a similar vein, the childbirth coefficients in the FD specification are positive for men, which indicates some short-run fatherhood premium, which would disappear in the long-run.

However, neither FD nor FE estimates correct for the possibility of poorer job matches for parents than for non-parents, which may help in explaining part of the observed wage differential. The comparison of column 2 (resp. 5) with column 3 (resp. 6) measures precisely the part of the gap that is explained by endogenous matching. Yet almost the whole motherhood pay gap remains – it is even slightly higher in absolute – while fatherhood penalties, if any, turn out to be no longer significant. To go further, we report several estimates of the childbirth coefficients in Tables 5 and 6 which correspond to different specifications of Equations (3) and (4). The coefficient of the third childbirth almost doubles when one omits to control for experience and for firm fixed effects. More generally, these Tables quantify the role played by each of the fourth explanations to the family gap mentioned in the literature.

The first explanation – human capital proxied by experience – would matter up to 50% of the adjusted motherhood pay gap, especially from the third childbirth onwards (comparison of columns 3 and 4c). Controlling for potential experience (column 4a) or neglecting to distinguishing among full-time and part-time experience (column 4b) biases the coefficients of interest. On the contrary, columns 3 and 4a look strikingly alike for both women and men. Disentangling part-time from full-time experience yields to smaller penalties.

The second explanation – individual unobserved heterogeneity – is not the major reason for lower hourly wages after childbirth. Columns 1 and 2 exhibit some differences but either they are not significant at 5%, or their magnitude is economically small. To a lesser extent, the rejection of this explanation holds also for men who experience some penalty after controlling for individual fixed effects.

The third explanation – firm matching – is rejected as far as women are concerned, which constitutes the main result of this paper: comparing estimates from columns 4c to those from column 5 does not permit to find out any significant difference. In other words, women who want to be mothers would not match firms which offer particularly lower wages. On the contrary, there is a small but significant difference in the case of men: omitting to control for firm fixed effects yields to small penalties while after including the latter, the effect of childbirth is not significant at usual levels. This result indicates that men who want to be fathers tend to select into firms that offer lower compensations; the negative coefficient in the FE specification accounts then for such a spurious correlation.

The fourth and remaining explanation for the residual motherhood penalty is discrimination, which could stem from a reallocation of work within firms or within establishments: after childbirth women would be confined to less risky missions and would hence receive less extra compensations.

To confirm those results, Tables 7 and 8 shed some light on the role played by both worker and firm unobserved heterogeneity. Table 7 displays the estimates of (5) which depicts the way estimated unobserved heterogeneity $\hat{\theta}_i$ vary with covariates. Once education has been controlled for, and after taking cohort effects into account, we do not find any significant difference between parents and non-parents in terms of individual average productivity. If anything, married men tend to receive slightly higher hourly wages. Table 8 presents correlations between individual and firm unobserved heterogeneity. Consistently with previous findings, this correlation is negative and amounts in our sample to -.21. It is approximately -.22 for non-mothers and -.26 for mothers, but -.2 for both fathers and non-fathers. These figures suggest that overall there are only limited firm matching forces that would trap parents into low wage firms, which does not exclude that some (high- or low- wage) individuals (parents or not) tend to match high- or low-wage firms. They also indicate that firm matching works slightly differently for women and for men since the previous correlation is higher for mothers than for non-mothers, while it is the same for fathers as for non-fathers.

Our results are also consistent with the literature focusing on the effect of marriage on wages. We estimate a "marriage pay premium" that amounts from 3.1% for men to 3.2% for women. Interestingly, most of the corresponding literature focused on men's marriage premium while we cannot reject that this premium is as high for women as it is for men.

Finally, the effect of part-time on hourly wages is positive and has already been discussed. Seniority and full-time experience exhibit diminishing marginal returns with peaks attained at respectively 6 and 14.7 years for women, against 6.7 and 12.9 years for men.

5.2 How much does the family gap contribute to the gender gap?

To evaluate the contribution of this gender-biased parenthood penalty to the gender gap, we simulate a counterfactual *scenario* in which women would experience the same childbirth penalty as men. Public interventions might well consist in promoting and facilitating paternity leaves, which could reduce or even eliminate such a difference. From women's observed wages w_{it}^o we compute therefore their simulated wages w_{it}^s in the case they actually face (non significant) fatherhood's penalties:

$$\log w_{it}^s = \log w_{it}^o + \sum_{k=1}^{5} (\beta_{ck}^{\text{Men}} - \beta_{ck}^{\text{Women}})\text{Childbirth}_{itk}. \tag{6}$$

We then estimate annual adjusted gender pay gaps Δ_t^o and Δ_t^s from both observed and simulated wages: denoting by G_i the gender dummy equal to 1 if individual i is a woman, $\forall m \in (o, s)$,

$$\log w_{it}^m = \Delta_t^m G_i + X_{it}'\beta^m + N_{jt}'\delta^m + \epsilon_{it}^m. \tag{7}$$

Figure 1 depicts the fraction of women's wages in terms of men's wages in both observed and counterfactual *scenarii*. Several remarks are worth being discussed. First, our sample is made up of individuals aged 26.9 on average for which the gender gap is low. In 2013, the French unadjusted gender gap is almost zero for individuals aged less than 25. Second, composition effects due to youth insertion in the labor market may explain the erosion of this gender gap during the 2000s. However, when the observed adjusted gender gap was about 17.5% in the mid-1990s, the counterfactual gender gap was still 17% –hence a small difference of hardly .5pp. On the contrary, at the end of the 2000s when the observed adjusted gender gap was less than 2% on this sample, women would experience almost the same wages as men in the counterfactual *scenario*. This convergence stems from the combination of two effects: the dynamics of the gender gap (which is *a priori* specific to our sample) and a larger contribution of the family gap to the gender gap (almost 2pp, which amounts quite to the whole gender gap).

5.3 Robustness checks

We proceed to two robustness checks of our results. First we examine the sensitivity of our estimates with respect to outliers. We perform several estimations with and without trimming hourly wages. Table 9 displays the results: column 1 corresponds to no trimming, column 2 corresponds to the elimination of hourly wages below 0.8 minimum hourly wage (which is also our base specification), column 3 to the elimination of hourly wages below 1 minimum hourly wage while column 4 further imposes a cap at 100 euros following Felfe (2012). Overall we find a limited impact on the motherhood penalty (eliminating outliers tends to reduce the estimated loss) while as far as men are concerned, the absence of trimming at the bottom of the distribution leads to significant and large fatherhood wage penalties; no trimming at the top results in close estimates.

Second we investigate whether different measures of experience alter our diagnosis. As already argued, when tackling the family gap issue seriously, it is important to compute the experience covariate as accurately as possible. Disposing of administrative data is an helpful tool that enables us to provide an almost ideal variable with little measurement error. It turns out that the definition of experience matters: on top of counting the amount of time spent on-the-job, distinguishing carefully among full-time and part-time experience has an impact of the estimated effect of children on wages. Childbirth coefficients differ slightly according to whether one controls for experience as a whole, or for both full-time and part-time experience. Potential experience (squared), which is a poor measure of the actual time spent in the labor workforce, does the worse job.

Third we check whether previous results are robust to the inclusion of occupational covariates in log hourly wage equations. In general we are reluctant to control for occupation in wage equations because it is likely to be correlated with unobserved determinants of wages including talent or productivity, hence occupation may be viewed as an endogenous variable. We check nevertheless whether controlling for such covariates alters dramatically our diagnosis, since there is no consensus in the literature on that topic. Table 11 displays the corresponding results and shows that not only signs and significance of childbirth effects remain once occupation (namely, dummies defined by the two-digit PCS-ESE French classification) has been controlled for, but also their magnitude. There is hardly an attenuation in the FE specification, but no significant difference is observed in the 2FE specification between columns 3 and 6 of Tables 4 and 11.

6 Conclusion

This paper has reexamined the family pay gap by resorting to linked employer-employee data and by controlling for three explanatory factors in wage equations: firm, worker and time. It provides a test of the firm matching explanation according to which endogenous selection of parents into low wage firms would explain the parenthood penalty. On French data over the 1995-2011 period, we estimate a linear model in the presence of two-way high dimensional fixed effects. We find a motherhood wage penalty of roughly -3% per child on the hourly wage. By contrast, fathers do not experience any significant loss consecutive to childbirths. There is little difference between estimates issued from specifications that control and that do not control for firm effects: hence we reject firm matching as the main explanation for gender-biased parenthood penalty. A remaining explanation is discrimination against mothers at work, which might stem from within-firm labor reallocation: mothers would be less exposed to risky missions, thus less likely to receive bonuses, or even trapped into low-wage trajectories. Testing explicitly for the presence of discrimination against mothers at work is a task that we leave for further research. Moreover, such a gender inequality is both unfair and inefficient, and legitimates further public intervention including campaigns against discrimination, the development of on-the-job childcare and the generalization (or/and extension) of the paternity leave.

References

Abowd, J. M., R. H. Creecy, and F. Kramarz (2002): "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data," Cornell University Working Paper.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High wage workers and high wage firms," *Econometrica*, 67(2), 251–333.

Abowd, J. M., F. Kramarz, and S. Woodcock (2008): "Econometric Analyses of Linked Employer-Employee Data," in *The Econometrics of Panel Data*, ed. by L. Matyas, and P. Sevestre, pp. 727–760. Springer.

Albrecht, J. W., P.-A. Edin, M. Sundström, and S. B. Vroman (1999): "Career interruptions and subsequent earnings: A reexamination using Swedish data," *Journal of Human Resources*, XXXIV(2), 294–311.

Beblo, M., S. Bender, and E. Wolf (2009): "Establishment-level wage effects of entering motherhood," *Oxford Economic Papers*, 61(suppl 1), i11–i34.

Becker, G. S. (1985): "Human capital, effort, and the sexual division of labor," *Journal of Labor Economics*, 3(1), S33–S58.

Booth, A. L., and M. Wood (2008): "Back-to-Front Down Under? Part-Time/Full-Time Wage Differentials in Australia," *Industrial Relations*, 47(1), 114–135.

Budig, M., and M. Hodges (2010): "Differences in disadvantage: How the wage penalty for motherhood varies across women's earnings distribution," *American Sociological Review*, 75(5), 705–728.

Budig, M. J., and P. England (2001): "The wage penalty for motherhood," *American Sociological Review*, 66(2), 204–225.

Buligescu, B., D. De Crombrugghe, G. Menteşoğlu, and R. Montizaan (2009): "Panel estimates of the wage penalty for maternal leave," *Oxford Economic Papers*, 61(suppl 1), i35–i55.

Davies, R., and G. Pierre (2005): "The family gap in pay in Europe: A cross-country study," *Labour Economics*, 12(4), 469–486.

England, P., J. Bearak, M. Budig, and M. Hodges (2013): "How the Motherhood Penalty Varies by Wage and Cognitive Skill: A Reassessment," Working paper.

Felfe, C. (2012): "The motherhood wage gap: What about job amenities?," *Labour Economics*, 19(1), 59–67.

Glauber, R. (2008): "Race and Gender in Families and at Work The Fatherhood Wage Premium," *Gender & Society*, 22(1), 8–30.

Godwin, W. (1798): *Memoirs of the Author of A Vindication of the Rights of Woman*. Broadview Press.

Joseph, O., A. Pailhé, I. Recotillet, and A. Solaz (2013): "The economic impact of taking short parental leave: Evaluation of a French reform," *Labour Economics*, 25, 63–75.

Killewald, A. (2013): "A Reconsideration of the Fatherhood Premium Marriage, Coresidence, Biology, and Fathers' Wages," *American Sociological Review*, 78(1), 96–116.

Kleven, H. J., C. Landais, and J. E. Søgaard (2015): "Parenthood and the Gender Gap: Evidence from Denmark," Working Paper.

Korenman, S., and D. Neumark (1992): "Marriage, Motherhood, and Wages," *Journal of Human Resources*, 27(2), 233–255.

Lequien, L. (2012): "The impact of parental leave duration on later career," Crest Working paper 2012-04.

Lundberg, S., and E. Rose (2000): "Parenthood and the earnings of married men and women," *Labour Economics*, 7(6), 689–710.

——— (2002): "The effects of sons and daughters on men's labor supply and wages," *Review of Economics and Statistics*, 84(2), 251–268.

Meurs, D., A. Pailhé, and S. Ponthieux (2010): "Child-related career interruptions and the gender wage gap in France," *Annales d'Économie et de Statistique*, 99/100, 15–46.

Meurs, D., and S. Ponthieux (2000): "Une mesure de la discrimination dans l'écart de salaire entre hommes et femmes," *Économie et statistique*, 337(1), 135–158.

Meurs, D., and S. Ponthieux (2014): "Gender Inequality," in *Handbook of Income Distribution SET vols. 2A-2B*, ed. by A. Atkinson, and F. Bourguignon. North Holland.

Miller, A. R. (2011): "The effects of motherhood timing on career path," *Journal of Population Economics*, 24(3), 1071–1100.

Mincer, J. (1958): "Investment in human capital and personal income distribution," *The Journal of Political Economy*, 66(4), 281–302.

Nielsen, H. S., M. Simonsen, and M. Verner (2004): "Does the Gap in Family-friendly Policies Drive the Family Gap?," *The Scandinavian Journal of Economics*, 106(4), 721–744.

OECD (2012): *Closing the Gender Gap: Act Now.* OECD Publishing.

Simonsen, M., and L. Skipper (2006): "The costs of motherhood: An analysis using matching estimators," *Journal of Applied Econometrics*, 21(7), 919–934.

——— (2012): "The family gap in wages: What wombmates reveal," *Labour Economics*, 19(1), 102–112.

Spence, M. (1973): "Job market signaling," *The Quarterly Journal of Economics*, 87(3), 355–374.

Waldfogel, J. (1997): "The effect of children on women's wages," *American Sociological Review*, 123, 209–217.

——— (1998): "The family gap for young women in the United States and Britain: Can maternity leave make a difference?," *Journal of Labor Economics*, 16(3), 505–545.

Wilde, E. T., L. Batchelder, and D. T. Ellwood (2010): "The mommy track divides: The impact of childbearing on wages of women of differing skill levels," NBER Working paper 16582.

Wollstonecraft, M. (1792): *A Vindication of the Rights of Woman.* Penguin UK.

Tables

Table 1: Sample - Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Woman	41531	0.480	0.500	0	1
First year in panel	41531	2003	4.689	1995	2011
Married	41531	0.189	0.391	0	1
One child	41531	0.165	0.372	0	1
Two children	41531	0.121	0.326	0	1
Three children	41531	0.039	0.194	0	1
Four children	41531	0.010	0.101	0	1
Five children	41531	0.004	0.064	0	1
Age	212189	26.900	7.378	16	65
Potential experience	212189	6.443	4.244	1	17
Full-time experience	212189	2.348	2.824	0	17
Part-time experience	212189	0.802	1.108	0	14
Seniority	212189	2.487	2.611	0	17
Nb. of working days	212189	237.2	132.1	1	360
Nb. of working hours	212189	1037	727	1	4400
Part time	212189	0.410	0.492	0	1
Net annual wage	212189	11047	12211	10.18	1067696
Net hourly wage	212189	9.98	7.36	3.51	1760
Seniority	301079	2.116	2.434	0.01	17

Sample of 41,531 French individuals working in the private sector from 1995 to 2011 (212,189 individual×year observations, 301,079 individual×firm×year observations). Wages: in 2011 euros. Full-time experience: in full-time full-year units.

Table 2: Estimates of part-time/full-time log hourly wage differential

	Women		Men	
Specification	Pooled OLS	FE	Pooled OLS	FE
(1)	-0.079***	-0.019***	-0.066***	-0.029***
	(0.002)	(0.003)	(0.002)	(0.003)
(2)	-0.052***	0.034***	-0.007***	0.065***
	(0.002)	(0.002)	(0.002)	(0.002)
(3)	-0.020***	0.036***	0.014***	0.062***
	(0.002)	(0.002)	(0.002)	(0.002)
(4)	-0.015***	0.036***	0.023***	0.061***
	(0.002)	(0.002)	(0.002)	(0.002)
(5a)	-0.006***	0.036***	0.029***	0.061***
	(0.002)	(0.002)	(0.002)	(0.002)
(5b)	0.008***	0.042***	0.044***	0.069***
	(0.002)	(0.002)	(0.002)	(0.002)
(5c)	0.038***	0.052***	0.064***	0.074***
	(0.002)	(0.002)	(0.002)	(0.002)
Observations	95499	95499	116690	116690

(1) contains a constant.
(2) adds worker characteristics (quadratic specification in age, dummy if married).
(3) adds firm characteristics (*département*, two-digit industry dummies and establishment size).
(4) adds a quadratic specification in seniority.
(5a) = (4) + potential experience (squared).
(5b) = (4) + experience (and its square).
(5c) = (4) + full-time, part-time experience (and their squares).

Table 3: Impact of marriages and childbirths on experience

	Women		Men	
	Full-time	Part-time	Full-time	Part-time
Marriage	0.310***	-0.019	0.596***	-0.046
	(0.051)	(0.027)	(0.051)	(0.024)
First childbirth	0.082*	0.153***	0.427***	0.004
	(0.038)	(0.020)	(0.035)	(0.016)
Second childbirth	-0.404***	0.427***	0.810***	0.014
	(0.067)	(0.035)	(0.059)	(0.027)
Third childbirth	-1.465***	0.443***	0.786***	0.111*
	(0.127)	(0.068)	(0.113)	(0.051)
Fourth childbirth	-2.716***	0.305*	0.746**	0.203
	(0.239)	(0.133)	(0.250)	(0.105)
Fifth childbirth	-3.581***	0.003	-0.753	0.232
	(0.525)	(0.161)	(0.426)	(0.205)
Year dummies	Yes	Yes	Yes	Yes
Individual effects	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes
Regional dummies	Yes	Yes	Yes	Yes
Observations	95499	95499	116690	116690
R^2	0.848	0.815	0.885	0.793

Clustered standard errors at the individual level in parentheses

Industry dummies: 39 two-digit dummies (NACE)

Table 4: Log hourly wages

	Women			Men		
	(1)	(2)	(3)	(4)	(5)	(6)
	FD	FE	2FE	FD	FE	2FE
Marriage	0.050***	0.034***	0.032***	0.066***	0.050***	0.031***
	(0.004)	(0.005)	(0.007)	(0.004)	(0.006)	(0.007)
First childbirth	-0.029***	-0.039***	-0.043***	0.018***	-0.000	0.004
	(0.003)	(0.004)	(0.005)	(0.003)	(0.004)	(0.005)
Second childbirth	-0.043***	-0.061***	-0.067***	0.026***	-0.021**	-0.003
	(0.005)	(0.006)	(0.008)	(0.005)	(0.006)	(0.008)
Third childbirth	-0.033***	-0.074***	-0.074***	0.023**	-0.060***	-0.016
	(0.008)	(0.011)	(0.015)	(0.008)	(0.011)	(0.013)
Fourth childbirth	-0.069***	-0.104***	-0.102***	0.034*	-0.053*	-0.020
	(0.016)	(0.021)	(0.030)	(0.016)	(0.021)	(0.028)
Fifth childbirth	-0.006	-0.061	-0.081	-0.059	-0.174***	-0.025
	(0.021)	(0.052)	(0.061)	(0.034)	(0.033)	(0.046)
Part-time	0.057***	0.058***	0.063***	0.074***	0.074***	0.084***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Seniority	0.019***	0.016***	0.012***	0.024***	0.018***	0.012***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Seniority^2 (1e-3)	-0.957***	-1.247***	-0.996***	-1.521***	-1.353***	-0.900***
	(0.112)	(0.114)	(0.133)	(0.109)	(0.135)	(0.159)
FT Experience	0.056***	0.039***	0.035***	0.062***	0.044***	0.041***
	(0.002)	(0.002)	(0.003)	(0.001)	(0.002)	(0.003)
FT Experience^2 (1e-3)	-1.576***	-1.367***	-1.188***	-1.938***	-1.841***	-1.589***
	(0.159)	(0.179)	(0.226)	(0.123)	(0.145)	(0.182)
PT Experience	0.028***	-0.009*	-0.002	0.046***	0.005	0.003
	(0.003)	(0.004)	(0.005)	(0.003)	(0.005)	(0.006)
PT Experience^2 (1e-3)	-2.424***	0.732	0.602	-3.540***	-0.644	0.786
	(0.442)	(0.526)	(0.709)	(0.512)	(0.847)	(0.938)
Year dummies	No	Yes	Yes	No	Yes	Yes
Individual effects	No	Yes	Yes	No	Yes	Yes
Industry dummies	Yes	Yes	No	Yes	Yes	No
Regional dummies	Yes	Yes	No	Yes	Yes	No
Firm size controls	Yes	Yes	No	Yes	Yes	No
Firm effects	No	No	Yes	No	No	Yes
Observations	63260	95499	135431	80808	116690	165648
Nb. individuals	15721	19932	19932	18012	21599	21599
Nb. firms	21513	31189	42937	25770	36408	49556
R^2	0.194	0.681	0.816	0.251	0.720	0.841

Clustered standard errors at the individual level in parentheses

Industry dummies: 39 two-digit dummies (NACE)

Firm size controls: 12 dummies

Table 5: Coefficients of childbirth in Mincer equations - Women

Specification	(1)	(2)	(3)	(4a)	(4b)	(4c)	(5)
First childbirth	-0.045***	-0.032***	-0.036***	-0.036***	-0.042***	-0.039***	-0.043***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Second childbirth	-0.069***	-0.072***	-0.073***	-0.073***	-0.069***	-0.061***	-0.067***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)
Third childbirth	-0.124***	-0.114***	-0.112***	-0.112***	-0.082***	-0.074***	-0.074***
	(0.010)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.015)
Fourth childbirth	-0.174***	-0.181***	-0.171***	-0.170***	-0.109***	-0.104***	-0.102***
	(0.013)	(0.025)	(0.025)	(0.025)	(0.024)	(0.024)	(0.030)
Fifth childbirth	-0.169***	-0.161**	-0.144**	-0.143**	-0.058	-0.061	-0.081
	(0.021)	(0.066)	(0.067)	(0.067)	(0.058)	(0.059)	(0.061)
Observations	95499	95499	95499	95499	95499	95499	135431
R^2	0.310	0.671	0.674	0.674	0.679	0.681	0.816

(1) = Pooled OLS with year and cohort dummies.
(2) = FE with year dummies.
(3) = (2) + a quadratic specification in seniority.
(4a) = (3) + potential experience.
(4b) = (3) + experience.
(4c) = (3) + full-time and part-time experience.
(5) = 2FE.

Table 6: Coefficients of childbirth in Mincer equations - Men

Specification	(1)	(2)	(3)	(4a)	(4b)	(4c)	(5)
First childbirth	0.002	0.010**	0.008*	0.010**	-0.000	-0.000	0.004
	(0.005)	(0.005)	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)
Second childbirth	0.010	-0.012*	-0.010	-0.007	-0.020***	-0.021***	-0.003
	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)
Third childbirth	-0.036**	-0.061***	-0.054***	-0.049***	-0.061***	-0.060***	-0.016
	(0.017)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Fourth childbirth	-0.006	-0.064***	-0.052**	-0.045*	-0.053**	-0.053**	-0.020
	(0.032)	(0.024)	(0.023)	(0.024)	(0.023)	(0.023)	(0.028)
Fifth childbirth	-0.161***	-0.211***	-0.199***	-0.194***	-0.172***	-0.174***	-0.025
	(0.044)	(0.043)	(0.042)	(0.042)	(0.038)	(0.037)	(0.046)
Observations	116690	116690	116690	116690	116690	116690	165648
R^2	0.337	0.713	0.716	0.716	0.720	0.720	0.841

(1) = Pooled OLS with year and cohort dummies.
(2) = FE with year dummies.
(3) = (2) + a quadratic specification in seniority.
(4a) = (3) + potential experience.
(4b) = (3) + experience.
(4c) = (3) + full-time and part-time experience.
(5) = 2FE.

Table 7: Individual unobserved heterogeneity - log hourly wages

	Women		Men	
	(1)	(2)	(3)	(4)
Married	-0.030***	-0.006	-0.048***	-0.016***
	(0.004)	(0.004)	(0.005)	(0.005)
One child	0.035***	0.004	0.030***	-0.006
	(0.004)	(0.004)	(0.005)	(0.005)
Two children	0.055***	0.011*	0.064***	0.010
	(0.005)	(0.005)	(0.006)	(0.006)
Three children	0.063***	0.006	0.078***	-0.002
	(0.008)	(0.008)	(0.010)	(0.010)
Four children	0.080***	0.011	0.116***	0.022
	(0.015)	(0.015)	(0.018)	(0.018)
Five children	0.106***	0.018	0.164***	0.039
	(0.021)	(0.020)	(0.035)	(0.034)
Education (highest degree)	Yes	Yes	Yes	Yes
Cohort effects	No	Yes	No	Yes
Observations	19932	19932	21599	21599
R^2	0.075	0.139	0.101	0.161

Note. The dependent variable is $\hat{\theta}_i$ (see Equation 5).

Table 8: Correlation between individual and firm unobserved heterogeneity

	Women	Men
No child	-0.222	-0.196
One child or more	-0.258	-0.200

Figure

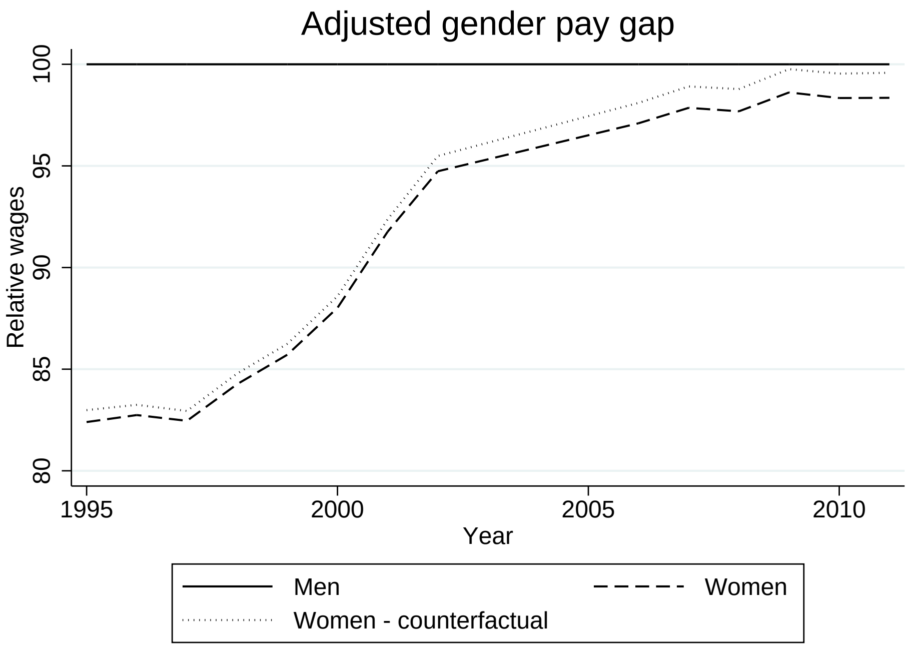

Figure 1: What if women experienced the same penalty as men regarding childbirth?

Robustness checks

Table 9: Sensitivity to the trimming of outliers - log hourly wages

	Women				Men			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	FE	FE	FE	FE	FE	FE	FE	FE
Marriage	0.025***	0.034***	0.033***	0.033***	0.023**	0.050***	0.050***	0.049***
	(0.006)	(0.005)	(0.005)	(0.005)	(0.007)	(0.006)	(0.006)	(0.006)
First childbirth	-0.045***	-0.039***	-0.031***	-0.031***	-0.017***	-0.000	-0.001	-0.001
	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)
Second childbirth	-0.083***	-0.061***	-0.049***	-0.049***	-0.067***	-0.021**	-0.014*	-0.015*
	(0.007)	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)
Third childbirth	-0.114***	-0.074***	-0.057***	-0.057***	-0.142***	-0.060***	-0.048***	-0.048***
	(0.013)	(0.011)	(0.011)	(0.011)	(0.013)	(0.011)	(0.011)	(0.011)
Fourth childbirth	-0.162***	-0.104***	-0.083***	-0.083***	-0.176***	-0.053*	-0.033	-0.031
	(0.027)	(0.021)	(0.020)	(0.020)	(0.024)	(0.021)	(0.021)	(0.021)
Fifth childbirth	-0.173*	-0.061	-0.050	-0.050	-0.295***	-0.174***	-0.143***	-0.134***
	(0.070)	(0.052)	(0.054)	(0.054)	(0.049)	(0.033)	(0.034)	(0.030)
Part-time	0.110***	0.058***	0.044***	0.043***	0.120***	0.074***	0.064***	0.063***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Seniority	0.016***	0.016***	0.017***	0.017***	0.017***	0.018***	0.019***	0.019***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Seniority^2 (1e-3)	-1.495***	-1.247***	-1.299***	-1.296***	-1.670***	-1.353***	-1.387***	-1.414***
	(0.129)	(0.114)	(0.110)	(0.110)	(0.152)	(0.135)	(0.130)	(0.125)
FT Experience	0.057***	0.039***	0.033***	0.033***	0.072***	0.044***	0.038***	0.038***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
FT Experience^2 (1e-3)	-2.943***	-1.367***	-1.114***	-1.080***	-3.715***	-1.841***	-1.558***	-1.566***
	(0.200)	(0.179)	(0.175)	(0.171)	(0.168)	(0.145)	(0.141)	(0.139)
PT Experience	-0.014**	-0.009*	-0.010**	-0.009*	0.028***	0.005	0.004	0.005
	(0.005)	(0.004)	(0.004)	(0.004)	(0.006)	(0.005)	(0.005)	(0.005)
PT Experience^2 (1e-3)	-0.172	0.732	0.894	0.792	-4.732***	-0.644	-0.574	-0.669
	(0.605)	(0.526)	(0.517)	(0.513)	(1.093)	(0.847)	(0.825)	(0.824)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm size controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	99967	95499	91825	91813	124977	116690	112560	112498
Nb. individuals	20300	19932	19665	19664	22213	21599	21288	21286
R^2	0.669	0.681	0.697	0.702	0.711	0.720	0.732	0.732

Same legend as Table 4.
Columns (1) and (5): no trimming.
Columns (2) and (6): base specification (hourly wage>.8 minimum wage).
Columns (3) and (7): hourly wage≥minimum hourly wage.
Columns (4) and (8): hourly wage∈ [minimum hourly wage; 100].

Table 10: Sensitivity to the specification of experience - log hourly wages

	Women				Men			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	FE	FE	FE	FE	FE	FE	FE	FE
Marriage	0.043***	0.043***	0.035***	0.034***	0.061***	0.062***	0.049***	0.050***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
First childbirth	-0.036***	-0.036***	-0.042***	-0.039***	0.008*	0.010*	-0.000	-0.000
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Second childbirth	-0.073***	-0.073***	-0.069***	-0.061***	-0.010	-0.007	-0.020**	-0.021**
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
Third childbirth	-0.112***	-0.112***	-0.082***	-0.074***	-0.054***	-0.049***	-0.061***	-0.060***
	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Fourth childbirth	-0.171***	-0.170***	-0.109***	-0.104***	-0.052*	-0.045*	-0.053*	-0.053*
	(0.022)	(0.022)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Fifth childbirth	-0.144*	-0.143*	-0.058	-0.061	-0.199***	-0.194***	-0.172***	-0.174***
	(0.060)	(0.060)	(0.051)	(0.052)	(0.038)	(0.038)	(0.034)	(0.033)
Part-time	0.040***	0.040***	0.046***	0.058***	0.058***	0.059***	0.067***	0.074***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Seniority	0.021***	0.021***	0.015***	0.016***	0.027***	0.025***	0.015***	0.018***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Seniority^2 (1e-3)	-1.330***	-1.323***	-1.329***	-1.247***	-1.782***	-1.609***	-1.127***	-1.353***
	(0.103)	(0.113)	(0.116)	(0.114)	(0.115)	(0.123)	(0.139)	(0.135)
Potential experience		0.031***				0.043***		
		(0.001)				(0.001)		
Pot. exp.^{2} (1e-3)		-0.012				-0.311***		
		(0.073)				(0.070)		
Experience			0.034***				0.051***	
			(0.002)				(0.002)	
Experience^2 (1e-3)			-0.648***				-1.898***	
			(0.148)				(0.130)	
FT Experience				0.039***				0.044***
				(0.002)				(0.002)
FT Experience^2 (1e-3)				-1.367***				-1.841***
				(0.179)				(0.145)
PT Experience				-0.009*				0.005
				(0.004)				(0.005)
PT Experience^2 (1e-3)				0.732				-0.644
				(0.526)				(0.847)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm size controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	95499	95499	95499	95499	116690	116690	116690	116690
Nb. individuals	19932	19932	19932	19932	21599	21599	21599	21599
R^2	0.674	0.674	0.679	0.681	0.716	0.716	0.720	0.720

Same legend as Table 4.

Table 11: Sensitivity to the inclusion of occupational covariates - log hourly wages

	Women			Men		
	(1)	(2)	(3)	(4)	(5)	(6)
	FD	FE	2FE	FD	FE	2FE
Marriage	0.036***	0.021***	0.022***	0.043***	0.030***	0.021**
	(0.004)	(0.005)	(0.006)	(0.004)	(0.005)	(0.007)
First childbirth	-0.022***	-0.033***	-0.040***	0.019***	0.005	0.007
	(0.003)	(0.004)	(0.005)	(0.003)	(0.003)	(0.005)
Second childbirth	-0.025***	-0.047***	-0.062***	0.029***	-0.005	0.004
	(0.004)	(0.005)	(0.008)	(0.005)	(0.006)	(0.007)
Third childbirth	-0.009	-0.059***	-0.069***	0.040***	-0.023*	0.002
	(0.007)	(0.010)	(0.014)	(0.008)	(0.010)	(0.012)
Fourth childbirth	-0.035*	-0.078***	-0.098***	0.052***	-0.009	0.008
	(0.015)	(0.019)	(0.029)	(0.014)	(0.019)	(0.027)
Fifth childbirth	0.021	-0.017	-0.055	-0.030	-0.105**	0.017
	(0.020)	(0.046)	(0.061)	(0.031)	(0.032)	(0.043)
Part-time	0.054***	0.055***	0.061***	0.070***	0.073***	0.083***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Seniority	0.017***	0.015***	0.011***	0.021***	0.018***	0.011***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Seniority^2 (1e-3)	-0.798***	-1.038***	-0.781***	-1.255***	-1.236***	-0.722***
	(0.105)	(0.102)	(0.125)	(0.100)	(0.118)	(0.143)
FT Experience	0.039***	0.021***	0.023***	0.045***	0.028***	0.029***
	(0.001)	(0.002)	(0.003)	(0.001)	(0.002)	(0.002)
FT Experience^2 (1e-3)	-0.763***	-0.436**	-0.625**	-1.112***	-0.949***	-1.068***
	(0.149)	(0.154)	(0.209)	(0.114)	(0.126)	(0.166)
PT Experience	0.026***	-0.010**	-0.006	0.040***	-0.000	-0.004
	(0.002)	(0.003)	(0.005)	(0.003)	(0.004)	(0.006)
PT Experience^2 (1e-3)	-2.092***	0.909	0.806	-3.211***	0.268	1.536
	(0.414)	(0.466)	(0.653)	(0.471)	(0.748)	(0.898)
Year dummies	No	Yes	Yes	No	Yes	Yes
Occupational dummies	Yes	Yes	Yes	Yes	Yes	Yes
Individual effects	No	Yes	Yes	No	Yes	Yes
Industry dummies	Yes	Yes	No	Yes	Yes	No
Regional dummies	Yes	Yes	No	Yes	Yes	No
Firm size controls	Yes	Yes	No	Yes	Yes	No
Firm effects	No	No	Yes	No	No	Yes
Observations	63260	95499	135431	80808	116690	165648
Nb. individuals	15721	19932	19932	18012	21599	21599
Nb. firms	21513	31189	42937	25770	36408	49556
R^2	0.294	0.720	0.829	0.368	0.764	0.854

Clustered standard errors at the individual level in parentheses

Industry dummies: 39 two-digit dummies (NACE)

Firm size controls: 12 dummies

Occupational dummies: 38 two-digit dummies (PCS-ESE)

A Appendix: data

Cleaning

We proceed to some cleaning of the DADS panel. First we recode the age variable as the difference between the current year and the year of birth. The former age variable exhibits some errors due to scan problems before the numerical DADS was introduced.[19] Second, *département* codes are sometimes one-digit instead of being two-digit; other *département* or region codes are missing. In that case we rely on other observations in the whole database in order to recover that information.

In the EDP database, we eliminate observations for which days or months of marriage or birth are equal either to 00 or 99, as well as observations for which the year of birth is 0000.

Selection

We restrict our attention to individuals born on October of even-numbered years: careers of individuals born on October of odd-numbered years is unknown before 2002. The most important selection is dictated by the necessity of measuring experience properly (see *infra*): we focus on individuals who entered the panel after 1995, which leaves us with 46,280 individuals (338,879 observations at the individual-year level and 489,852 observations at the individual-firm-year level). We eliminate further individuals whose net annual earnings are missing or less than 10 euros in 2011 terms. We also restrict our sample to individuals aged 16 to 65, working at least 10 hours a year, whose job duration is consistent with worked hours (for instance, the ratio of the latter over the former must be less than 24), which leaves us with 45,483 individuals (317,476 individual-year observations). After trimming observations with a hourly wage that is smaller than 80% of the legal minimum wage,[20] and after dropping years 2003 to 2005, our estimation sample is composed of 41,531 individuals (212,189 individual-year observations and 301,079 individual-firm-year observations). Among those individuals, 19,932 are women while 21,599 are men. Last but not least, we define time-varying variables for marriage (parenthood) as the fact of being married (experiencing a childbirth)

[19] Such variations may spuriously "identify" the slope of age in a FE specification as in Lequien (2012) where both individual and year effects are included in a linear fashion.

[20] We proceed to robustness checks with respect to the 80% threshold in Section 5.3.

before time t for individual i.

Definition of main employment

Aggregating data at the individual-year level requires to define for each individual her main employment in the year. We select the employment with (in successive order) the highest number of working days, the highest wage, a full-time position (if any) and the highest number of worked hours. If there are still ties after applying those criteria, we choose the job with the last SIREN in lexicographical order –to keep the code deterministic. Finally, if several observations resisted to the last iteration, we would consider them as authentic doubles and eliminate them –which does not happen here. We define job characteristics (private/public sector, industry, geographic location, firm's size, full-time/part-time, but also seniority) at the individual-year level as being related to the main employment. We sum wages and working hours, and define working days as the minimum of 360 (the annual number of working days in the DADS by convention) and the sum of working days over the whole year.

Computation of experience

Mincer (1958) demonstrated how important it is to control properly for experience and seniority in wage equations. We devote much attention to compute these variables as precisely as possible. Seniority is defined as the difference between the current date and the first appearance of a pair (individual, firm). Thanks to the comprehensive nature of the DADS panel, it is possible to reconstitute the whole salaried career of an individual, hence to compute his experience from observed working times. Experience will thus be defined as closely as possible as the amount of salaried time spent on the labor market. Since worked hours have been available from 1995 onwards only, we restrict our attention to individuals who entered the panel after 1995. We consider that workers increase their experience variable every year by their share of working hours.