conseil d'analyse économique

Reducing the Gender Wage Gap

Methodological Appendix

Antoine Bozio, Brigitte Dormont and Cecilia García-Peñalosa

Scientific Advisor: Manon Domingues Dos Santos

October 2014

Decomposition of the wage gap in 2012

Following on from the work of Blinder (1973) and Oaxaca (1973), the wage gap between demographic groups is usually broken down into two components. The first component, referred to as the "explained component", is due to differences in observable individual characteristics (level of qualification, work experience, job type…). The "unexplained component" arises from the individual characteristics of the two demographic groups being assessed differently. However one should remember, when estimating, to take into account the behaviour-related potential selection biases (see Heckman, 1979). Indeed, by definition, wages can only be observed for those that are employed. Yet expectations of a high salary are a likely determinant of the labour supply and, conversely, inadequate characteristics could lead to exclusion from the labour market. It is therefore the same individual characteristics that influence both employment status and wage. One way to deal with this simultaneity bias is to adopt, as Meurs and Ponthieux (2006) do, a two-step approach: at first, the status of the individual (employed or not) is regressed on his/her characteristics (education, number of children, etc.); then the individual's wage is regressed on these same characteristics, controlling for the person's probability to be employed.

The wage equations for women and men are respectively written: $W_i^F = \beta^F X_i + \tau^F \lambda_i + u_i$ et $W_i^H = \beta^H X_i + \tau^H \lambda_i + v_i$, where W_i^F and W_i^H represent the logarithm of women and men's wages, X_i the individual characteristics, β^F and β^H the coefficients for women and men and u_i , v_i are error terms (with a mean of zero), the index *i* denoting the individual. The term λ_i is the Mills ratio, computed based on the equation estimated during the first stage, and which corrects the selection bias. We also estimate an equation of uniform wage for women and for men, $W_i = \beta X_i + \tau \lambda_i + \varepsilon_i$, where the β coefficient can be interpreted as the "reference yield" of individual characteristics.

The wage gap between men and women is then decomposed according to the Oaxaca and Ramsom method (1999): the difference in average wages is written:

$$\overline{W}^H - \overline{W}^F = \hat{\beta}\left(\overline{X}^H - \overline{X}^F\right) + \left[\left(\hat{\beta}^H - \hat{\beta}\right)\overline{X}^H - \left(\hat{\beta}^F - \hat{\beta}\right)\overline{X}^F\right] + \left(\hat{\tau}^H \overline{\lambda}^H - \hat{\tau}^F \overline{\lambda}^F\right)$$

where each $\hat{\beta}$ represents the estimated value of the coefficient, and $\overline{X}$ is the mean of X.

The first term, $\hat{\beta}\left(\overline{X}^H - \overline{X}^F\right)$, is the wage gap attributable to differences between genders in

average characteristics. These characteristics are valued using coefficients estimated on the whole population: if women have a higher level of education than men, it contributes positively to their wage relative to that of men. This first term is the "explained component" of the wage gap.

The second term, $\left[\left(\hat{\beta}^H - \hat{\beta}\right)\overline{X}^H - \left(\hat{\beta}^F - \hat{\beta}\right)\overline{X}^F\right]$, represents the gap in the value placed on the characteristics of men and women: given the same level of education, if the labour market places less value on the skills of women compared to those of men, this effect contributes negatively to the wage of women relative to men. This second term is the "unexplained component" of the wage gap.

The last term, $\left(\hat{\tau}^H \overline{\lambda}^H - \hat{\tau}^F \overline{\lambda}^F\right)$, corrects for the fact that employed women and men do not have the same characteristics as unemployed or inactive women and men: if a limited wage perspective more heavily discourages women, than it does men, from participating in the labour market (or if it further lowers their chances of securing employment), then this selection effect will contribute positively to women's wages relative to those of men (because the women in the sample are those who, on average, are eligible for a higher salary).

The estimates computed here mobilize data from the 2012 Employment Survey *Emploi en continu* (INSEE) related to individuals being questioned in the first wave. The explanatory variables are listed in Tables 1 (selection equation) and 2 (wage equation). They may differ from those used by Meurs and Ponthieux (2006) due to modifications made to such Employment Surveys from 2003 onwards. The reference population is the working-age population, that is, individuals aged 16-65. For the selection equation, we excluded those individuals who are not available to take up paid employment (students, pensioners, self-employed), leaving us with employees, the "purely inactive" and the unemployed. For the wage equation, we excluded employees whose usual weekly working hours amount to fewer than 10, as well as those whose statutes place them between studies and employment (apprentices and trainees in vocational education streams).

1. Selection equation: explanatory variables for paid employment

Explanatory variable	Content	
Human capital	Age; age squared	
	Years of study	
Family situation	6 situations	Partnered or not; with children or not; fewer than 2 children, 3 children or more
Labour status in year $n-1$	4 statuses	Employed; unemployed; studying; inactive
Region of residence	22 indicators	
Size of urban unit	5 indicators	
Owner-occupied dwelling	2 indicators	Yes; no
Birth country	2 indicators	France; other

2. Wage equation: explanatory variables for the logarithm of monthly wage

Explanatory variable	Content	
Human capital	Education Potential work experience; potential work experience squared Seniority in the firm; Seniority in the firm squared	5 levels Age minus years of study Number of years
Job characteristics	Number of hours worked per week Weekly hours in the contract Unskilled position Type of contract Employment sector Job category Function Sector of activity Specificity of the position	 Full time; long-hours, average-hours, short-hours part time Fixed-term contract; other Public; private 5 indicators 10 indicators 9 indicators Working on saturdays, sundays, at night
Birth country Residency Selection effect	France; other Paris and its surroundings; other Mills ratio	

Tables 3 and 4 show the estimation results for the year 2012. The results of the selection equations (Table 3) are standard and similar to those of Meurs and Ponthieux (2006): all other things being equal, the probability of being employed is higher for men than for women; this probability decreases with the number of children for women only. For wage equations (Table 4), the most striking result is that the selection effect switches signs in 2012 compared to what was observed for the years 1990 and 2002: in 2012, the selection effect pushes women's wages upwards compared to men's wages. This same result was highlighted for the United States (see e.g. Mulligan and Rubinstein, 2008).

3. Selection equations
Dependent variable: the individual is employed

	Women		Men	
	Coefficient	Chi2 Wald	Coefficient	Chi2 Wald
Constant	– 0.0180	33.68	0.2302	5337.89
Age	0.0732	208860.41	0.0567	123119.54
Age squared	– 0.00083	180994.18	– 0.00068	117973.49
Years of study	0.0180	52768.04	0.0274	111169.35
Family situation:				
• Single	*Réf.*		*Réf.*	
• Single parent	– 0.2187	76517.59	– 0.2378	62824.45
• Partnered, no children	– 0.0584	7616.8947	0.1733	60257.02
• Partnered, 2 children or less	– 0.0979	20586.18	0.1695	67976.49
• Partnered, 3 children or more	– 0.2850	69792.43	0.0539	2485.24
• Other	– 0.1707	20896.27	– 0.00892	55.34
Labour status in *n* – 1				
• Employed	*Réf.*		*Réf.*	
• Unemployed	– 1.8918	8477569.51	– 1.8791	8388758.38
• Studying	– 0.9146	800437.80	– 1.2833	1449214.90
• Inactive	– 2.9315	16086180.7	– 2.9170	5755160.87
Owner-occupied dwelling	0.2042	160947.57	0.2455	209803.41
Born abroad	– 0.2607	156088.35	– 0.2469	126708.42
% conformity	91.5		88.6	
Adjusted pseudo-R2	0.503		0.415	
Number of observations	20 664		17 883	

Interpretation: estimation method = probit

Source: authors' computations

4. Gains equations – coefficients for the main variables
Dependent variable: monthly wage in euro, 2012

	Women		Men	
	Coefficient	Standard deviation	Coefficient	Standard deviation
Education: BEP, CAP	0.052	0.0101	0.075	0.0081
Education: Baccalaureate	0.078	0.0115	0.125	0.0110
Education: 2 years of undergraduate studies	0.185	0.0129	0.174	0.0129
Education: highest diploma	0.207	0.0145	0.288	0.0162
Potential work experience	0.007	0.0011	0.015	0.0011
Potential work experience, squared	$-1.29*10^{-4}$	$2.18*10^{-5}$	$-2.399*10^{-4}$	$2.29*10^{-5}$
Seniority in the firm	0.011	0.0010	0.010	0.0012
Seniority in the firm, squared	$-7.96*10^{-5}$	$2.53*10^{-5}$	$-8.39*10^{-5}$	$2.91*10^{-5}$
Number of weekly hours worked	0.500	0.0302	0.468	0.0263
Long-hours part time	– 0.106	0.0104	– 0.254	0.0321
Average-hours part time	– 0.227	0.0179	– 0.375	0.0264
Short-hours part time	– 0.514	0.0461	– 0.662	0.0593
Unskilled position	– 0.521	0.0162	– 0.539	0.0168
Public sector	0.060	0.0085	0.014	0.0168
Fixed-term contract	– 0.116	0.0113	– 0.127	0.0165
Born abroad	0.08	0.0080	0.07	0.0087
Living in Paris and its surroundings	0.01	0.0102	0.01	0.0110
Inverse of Mills ratio (selection)	– 0.04	0.0092	– 0.01	0.0065
R2	0.697		0.632	
Number of observations	13 837		13 638	

Interpretation: The regressions also included job category, function, sector of activity, and the specificities of the job. Estimation method = ordinary least squares

Source: authors' computations

We can then decompose the 2012 average gender wage gap using data and econometric estimates, with results being shown in Table 5. The "explained component" of the wage gap (differences in levels of education, work experience, position held, hours worked) accounts for nearly 72% of the wage gap, while the "unexplained component" (individual characteristics being assessed differently) amount to just over a quarter, with the remainder being due to the selection effect.

5. Decomposition of the average monthly wage gap

	1990		2002		2012	
	Gap	% of total	Gap	% of total	Gap	% of total
Wage gap in %	22.97	—	22.28	—	24.50	—
Explained component	0.186	71.3	0.192	76.2	0.201	71.6
• Education	– 0.013	– 5.0	– 0.015	– 6.0	– 0.011	– 4.1
• Work experience	0.001	0.4	0.001	0.4	0.003	1.1
• Employment structure	0.076	29.1	0.085	33.7	0.085	30.2
• Hours worked	0.117	44.8	0.121	48.0	0.124	44.3
Unexplained component	0.079	30.3	0.069	27.4	0.072	25.6
Selection effect	– 0.004	– 1.5	– 0.006	– 2.4	0.008	2.8
Total	0.261	100	0.252	100	0.281	100

Interpretation: In 2012, the "unexplained component" of the wage gap, in logarithm, was of 0.072, which amounted to 25.6% of the total gap (0.072/0.281=25.6%). Based on a wage gap of 24.5%, the contribution of this component is therefore of 6.3 percentage points (24.5*0.256).

Source: See Meurs and Ponthieux (2006) for 1990 and 2002. Authors' computations for 2012, based on regressions from Table 4.

Of particular relevance within the "explained component" of the wage gap is the disparity in positions held. Table 6 shows that the sectoral segregation of women is not particularly strong in France compared to other countries; however, this must be weighed against a relatively high female participation rate, which should go hand in hand with the greater diversification of positions held.

6. Sectoral segregation and difference in labour market participation between men and women in 2010

	Segregation index	Gender participation gap
Belgium	0.916	13.08
Denmark	1.049	9.27
Finland	1.11	5.64
France	0.89	10.37
Greece	0.93	20.07
Ireland	0.94	15.84
Italy	0.74	21.18
Luxembourg	1.11	16.05
Norway	1.02	6.30
Poland	0.95	15.84
Spain	0.97	15.82
Sweden	1.01	6.55
Switzerland	0.83	14.47
Turkey	0.79	43.16
United Kingdom	0.96	13.02

Interpretation: Put forward by Charles and Grusky (1995), the gender segregation index *R* measures the gap between the distribution of occupations held as observed today and the distribution that would prevail, were there to be a proportional representation of genders in each category. Let M_i be the number of men employed in occupation *i*, F_i the number of women in this same occupation and *I* the number of occupation categories. The index is written:

$$R = \frac{1}{I}\sum_{i=1}^{I}\left|Ln\left(\frac{F_i}{M_i}\right) - \left[\frac{1}{I}\sum_{i=1}^{I}Ln\left(\frac{F_i}{M_i}\right)\right]\right|$$

In a perfectly integrated labour market, *R* is equal to 0.

Source: Authors' computations based on ILOSTAT, which splits occupations into seven categories.

Bibliography

Blinder A.S. (1973): « Wage Discrimination: Reduced Form and Structural Estimates », *The Journal of Human Ressources*, vol. 8, no 4, pp. 436-455.

Charles M. and Grusky D. (1995): « Models for Describing the Underlying Structure of Sex Segregation », *American Journal of Sociology*, vol. 100, no 4, pp. 931-71

Heckman J. (1979): « Sample Selection Bias as a Specification Error », *Econometrica*, vol. 47, no 1, pp. 153-161.

Meurs D. and S. Ponthieux (2006): « L'écart de salaire entre les hommes et les femmes peut-il encore baisser », *Économie et Statistique*, no 398-399, pp. 99-130.

Mulligan C.B. and Y. Rubinstein (2008): « Selection, Investment, and Women's Relative Wages Over Time », *Quarterly Journal of Economics*, vol. 123, no 3, pp. 1061-1110.

Oaxaca R. and M. Ransom (1999): « Identification in Detailed Wage Decompositions », *The Review of Economics and Statistics*, vol. 81, no 1, pp. 154-157.

Oaxaca R.L. (1973): « Male-Female Wage Differential in the Urban Labor Markets », *International Economic Review*, vol. 14, no 3, pp. 693-709.