



# Réduire les inégalités de salaires entre femmes et hommes

## Annexe méthodologique

Antoine Bozio, Brigitte Dormont et Cecilia García-Peñalosa

Conseiller scientifique : Manon Domingues Dos Santos

*Octobre 2014*

### Décomposition de l'écart de salaires en 2012

Depuis les travaux de Blinder (1973) et Oaxaca (1973), les écarts de salaires entre deux groupes démographiques sont usuellement décomposés en deux parties. Une première composante, dite « part expliquée », est imputable aux différences de caractéristiques individuelles observables (niveau de qualification, expérience professionnelle, type d'emploi...). La seconde composante, dite « part inexpliquée », correspond aux différences de valorisation des caractéristiques individuelles retenues entre les deux groupes démographiques. Il convient toutefois de prendre en considération, lors des estimations, les éventuels biais de sélection liés aux comportements d'activité (voir Heckman, 1979). En effet, le salaire n'est par définition observé que pour les personnes en emploi. Or l'anticipation d'un salaire élevé est un déterminant probable de l'offre de travail. À l'inverse, les personnes éloignées de l'emploi le sont peut-être en raison de caractéristiques inadéquates. Les mêmes caractéristiques individuelles influencent donc à la fois le statut vis-à-vis de l'emploi et le salaire en cas d'emploi. Une façon de traiter ce biais de simultanéité est d'adopter, comme Meurs et Ponthieux (2006)<sup>(1)</sup>, une démarche en deux étapes : dans un premier temps, on régresse le statut de l'individu (en emploi ou non) sur ses caractéristiques (éducation, nombre d'enfants, etc.) ; puis on régresse le salaire individuel sur ces mêmes caractéristiques, en contrôlant pour la probabilité de chacun d'être en emploi.

Les équations de salaires pour les femmes et les hommes sont respectivement  $W_i^F = \beta^F X_i + \tau^F \lambda_i + u_i$  et  $W_i^H = \beta^H X_i + \tau^H \lambda_i + v_i$ , où  $W_i^F$  et  $W_i^H$  représentent le logarithme

---

Nous remercions Laurence Bouvard et Maty Konte pour leur excellente assistance, ainsi que Dominique Meurs et Sophie Ponthieux pour nous avoir transmis leurs programmes d'estimation.

(1) Voir en particulier les encadrés 2 et 3. Nous adoptons une démarche similaire et étendons l'étude à l'année 2012.

des salaires des femmes et des hommes,  $X_i$  les caractéristiques individuelles,  $\beta^F$  et  $\beta^H$  les coefficients pour les femmes et les hommes et  $u_i$ ,  $v_i$  sont les termes d'erreur (de moyenne nulle), l'indice  $i$  désignant l'individu. Le terme  $\lambda_i$  est le ratio de Mills, calculé à partir de l'équation estimée en première étape, qui corrige le biais de sélection. Nous estimons aussi une équation de salaire uniforme pour les femmes et pour les hommes,  $W_i = \beta X_i + \tau \lambda_i + \varepsilon_i$ , dans laquelle le coefficient  $\beta$  peut être interprété comme le « rendement de référence » des caractéristiques individuelles.

L'écart salarial entre hommes et femmes est alors décomposé selon la méthode d'Oaxaca et Ramsom (1999) : la différence des salaires moyens s'écrit :

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

où chaque  $\hat{\beta}$  représente la valeur estimée du coefficient  $\bar{X}$  est la moyenne de  $X$ .

Le premier terme,  $\hat{\beta}(\bar{X}^H - \bar{X}^F)$ , est l'écart de salaire imputable aux différences de caractéristiques moyennes entre hommes et femmes, valorisées à l'aide des coefficients estimés sur toute la population : si les femmes ont un niveau d'éducation supérieur aux hommes, cela contribue positivement à leur salaire relativement aux hommes. Ce premier terme est la *part « expliquée »* des différences salariales.

Le deuxième terme,  $[(\hat{\beta}^H - \hat{\beta})\bar{X}^H - (\hat{\beta}^F - \hat{\beta})\bar{X}^F]$ , représente la différence de valorisation entre les caractéristiques des hommes et celles des femmes : à niveau d'éducation identique, si le marché du travail valorise moins les compétences des femmes que celles des hommes, cet effet contribue négativement au salaire des femmes relativement à celui des hommes. Ce deuxième terme est la *part « non expliquée »* de l'écart salarial.

Le dernier terme,  $(\hat{\tau}^H \bar{\lambda}^H - \hat{\tau}^F \bar{\lambda}^F)$ , corrige du fait que les femmes et les hommes en emploi n'ont pas les mêmes caractéristiques que les femmes et les hommes au chômage ou en inactivité : si une perspective salariale limitée décourage davantage les femmes que les hommes à participer au marché (ou réduit davantage leur accès à l'emploi), cet *effet de sélection* contribue positivement au salaire des femmes relativement à celui des hommes (car les femmes de l'échantillon sont celles qui, en moyenne, peuvent prétendre à un salaire plus élevé).

Les estimations mobilisent les données de l'enquête *Emploi en continu* de 2012 (INSEE) relatives aux individus en première interrogation. Les variables explicatives sont répertoriées dans les tableaux 1 (équation de sélection) et 2 (équation de salaire). Elles diffèrent parfois de celles utilisées par Meurs et Ponthieux (2006) en raison des changements survenus dans les enquêtes *Emploi* à partir de 2003. La population de référence est celle des individus en âge de travailler, âgés de 16 à 65 ans. Pour l'équation de sélection, nous avons exclu les personnes qui ne sont pas disponibles pour occuper un emploi salarié (étudiants, retraités, indépendants). Restent donc les salariés, les « inactifs purs » et les chômeurs. Pour l'équation de salaire, nous avons exclu les salariés dont l'horaire habituel hebdomadaire est inférieur à 10 heures ainsi que les statuts à la limite des études et de l'emploi (apprentis et stagiaires de la formation professionnelle).

## 1. Équation de sélection : variables explicatives de l'emploi salarié

Variable explicative	Contenu	
Capital humain	Âge et son carré Nombre d'années d'études	
Situation familiale	6 situations	couple ou non, avec ou sans enfant, jusqu'à 2 enfants, 3 enfants ou plus
Statut d'activité l'année $n - 1$	4 statuts	emploi, chômage, étude, inactivité
Région d'habitation	22 indicatrices	
Taille de l'unité urbaine	5 indicatrices	
Propriétaire du logement occupé	2 indicatrices	oui, non
Pays de naissance	2 indicatrices	France, autre

## 2. Équation de salaire : variables explicatives du logarithme du salaire mensuel

Variable explicative	Contenu	
Capital humain	niveau d'éducation expérience potentielle et son carré ancienneté dans l'entreprise et son carré	5 niveaux âge moins nombre d'années d'études en années
Caractéristiques des emplois	nombre d'heures par semaine type d'horaire  poste non qualifié type de contrat secteur d'emploi catégorie professionnelle fonction secteur d'activité particularités du poste	temps complet, temps partiel long, moyen, court  durée déterminée, autre public, privé 5 niveaux 10 niveaux 9 niveaux travail samedi, dimanche, nuit
Pays de naissance	France, autre	
Résidence	en Région parisienne, autre	
Effet de sélection	ratio de Mills	

Les tableaux 3 et 4 présentent les résultats des estimations pour l'année 2012<sup>(2)</sup>. Les résultats des équations de sélection (tableau 3) sont standards et proches de ceux de Meurs et Ponthieux (2006) : toutes choses égales par ailleurs, la probabilité d'être en emploi est plus élevée pour les hommes que pour les femmes ; elle décroît avec le nombre d'enfants uniquement pour les femmes. Pour les équations de salaires (tableau 4), le résultat le plus marquant est le fait que l'effet de sélection change de signe en 2012 par rapport à ce qu'on obtient sur les années 1990 et 2002 : en 2012, l'effet de sélection pousse les salaires des femmes vers le haut par rapport à celui des hommes. Ce même résultat a été mis en exergue pour les États-Unis (voir, par exemple, Mulligan et Rubinstein, 2008).

(2) Pour les résultats relatifs à 1990 et 2002, voir Meurs et Ponthieux (2006) *op.cit.*

### 3. Équations de sélection

Variable dépendante : l'individu est en emploi salarié

	Femmes		Hommes	
	Coefficient	Chi2 Wald	Coefficient	Chi2 Wald
Constante	- 0,0180	33,68	0,2302	5337,89
Âge	0,0732	208860,41	0,0567	123119,54
Âge <sup>2</sup>	- 0,00083	180994,18	- 0,00068	117973,49
Nombre d'années d'études	0,0180	52768,04	0,0274	111169,35
Situation familiale				
• personne seule	<i>Réf.</i>		<i>Réf.</i>	
• chef de famille monoparentale	- 0,2187	76517,59	- 0,2378	62824,45
• en couple sans enfant	- 0,0584	7616,8947	0,1733	60257,02
• en couple avec 2 enfants au plus	- 0,0979	20586,18	0,1695	67976,49
• en couple avec 3 enfants et plus	- 0,2850	69792,43	0,0539	2485,24
• autres	- 0,1707	20896,27	- 0,00892	55,34
Statut d'activité en $n - 1$				
• emploi	<i>Réf.</i>		<i>Réf.</i>	
• chômage	- 1,8918	8477569,51	- 1,8791	8388758,38
• études	- 0,9146	800437,80	- 1,2833	1449214,90
• inactivité	- 2,9315	16086180,7	- 2,9170	5755160,87
Propriétaire du logement occupé	0,2042	160947,57	0,2455	209803,41
Né à l'étranger	- 0,2607	156088,35	- 0,2469	126708,42
% concordant	91,5		88,6	
Pseudo-R2 ajusté	0,503		0,415	
Nombre d'observations	20 664		17 883	

Lecture : Méthode d'estimation = régression probit.

Source : Calculs des auteurs.

### 4. Équations de gains – Coefficients des principales variables

Variable dépendante : salaire mensuel en euro, 2012

	Femmes		Hommes	
	Coefficient	Écart type	Coefficient	Écart type
Éducation : BEP, CAP	0,052	0,0101	0,075	0,0081
Éducation : Bac	0,078	0,0115	0,125	0,0110
Éducation : Bac +2	0,185	0,0129	0,174	0,0129
Éducation : diplôme le plus élevé	0,207	0,0145	0,288	0,0162
Expérience potentielle	0,007	0,0011	0,015	0,0011
Expérience potentielle au carré	- 1,29*10 <sup>-4</sup>	2,18*10 <sup>-5</sup>	- 2,399*10 <sup>-4</sup>	2,29*10 <sup>-5</sup>
Ancienneté dans l'entreprise	0,011	0,0010	0,010	0,0012
Ancienneté dans l'entreprise au carré	- 7,96*10 <sup>-5</sup>	2,53*10 <sup>-5</sup>	- 8,39*10 <sup>-5</sup>	2,91*10 <sup>-5</sup>
Nombre d'heures par semaine	0,500	0,0302	0,468	0,0263
Temps partiel long	- 0,106	0,0104	- 0,254	0,0321
Temps partiel moyen	- 0,227	0,0179	- 0,375	0,0264
Temps partiel court	- 0,514	0,0461	- 0,662	0,0593
Poste non qualifié	- 0,521	0,0162	- 0,539	0,0168
Secteur public	0,060	0,0085	0,014	0,0168
CDD	- 0,116	0,0113	- 0,127	0,0165
Né à l'étranger	0,08	0,0080	0,07	0,0087
Résidence en Région parisienne	0,01	0,0102	0,01	0,0110
Inverse du ratio de Mills (sélection)	- 0,04	0,0092	- 0,01	0,0065
R2	0,697		0,632	
Nombre d'observations	13 837		13 638	

Lecture : Les régressions incluent aussi la catégorie professionnelle, la fonction, le secteur d'activité, et les particularités du poste. Méthode d'estimation : moindres carrés ordinaires.

Source : Calculs des auteurs.

On peut alors décomposer l'écart salarial moyen entre hommes et femmes en 2012 à l'aide des données et des estimations économétriques. Les résultats sont présentés dans le tableau 5. La part « expliquée » de l'écart salarial (différences de niveaux d'éducation, d'expérience, d'emplois occupés, de durée du travail) représente près de 72 % de l'écart total ; la part « non expliquée » (différence de valorisation des caractéristiques individuelles) représente un peu plus du quart ; le reste est dû à l'effet de sélection.

## 5. Décomposition de l'écart de salaires mensuels moyens

	1990		2002		2012	
	Écart	% du total	Écart	% du total	Écart	% du total
Écart des salaires en %	22,97	—	22,28	—	24,50	—
Part « expliquée »	0,186	71,3	0,192	76,2	0,201	71,6
• éducation	- 0,013	- 5,0	- 0,015	- 6,0	- 0,011	- 4,1
• expérience	0,001	0,4	0,001	0,4	0,003	1,1
• structure des emplois	0,076	29,1	0,085	33,7	0,085	30,2
• durée de travail	0,117	44,8	0,121	48,0	0,124	44,3
Part « non expliquée »	0,079	30,3	0,069	27,4	0,072	25,6
Effet de sélection	- 0,004	- 1,5	- 0,006	- 2,4	0,008	2,8
Total	0,261	100	0,252	100	0,281	100

*Lecture :* En 2012, la part « non expliquée » de l'écart salarial, en logarithmes, est de 0,072, ce qui représente 25,6% de l'écart total ( $0,072/0,281 = 25,6\%$ ). Sur un écart de salaire de 24,5 %, la contribution de cette part est de 6,3 points de pourcentage ( $24,5 \times 0,256$ ).

*Sources :* Cf. Meurs et Ponthieux (2006) pour 1990 et 2002. Calculs des auteurs pour 2012, à partir des régressions dans le tableau 5.

Au sein de la part « expliquée » de l'écart salarial, la différence des postes occupés tient une place importante. Le tableau 6 montre que la ségrégation sectorielle des femmes n'est pas particulièrement forte en France par rapport à d'autres pays, mais ceci doit être mis en regard d'un taux d'activité féminin relativement élevé qui devrait aller de pair avec une diversification plus forte des emplois occupés.

## 6. Ségrégation sectorielle et différence de participation sur le marché du travail entre hommes et femmes pour l'année 2010

	Indice de ségrégation	Écart de participation H-F
Belgique	0,916	13,08
Danemark	1,049	9,27
Espagne	0,97	15,82
Finlande	1,11	5,64
France	0,89	10,37
Grèce	0,93	20,07
Irlande	0,94	15,84
Italie	0,74	21,18
Luxembourg	1,11	16,05
Norvège	1,02	6,30
Pologne	0,95	15,84
Royaume-Uni	0,96	13,02
Suède	1,01	6,55
Suisse	0,83	14,47
Turquie	0,79	43,16

*Lecture* : L'indice de ségrégation entre hommes et femmes,  $R$ , a été proposé par Charles et Grusky (1995). Soit  $M_i$  le nombre d'hommes employés dans le poste  $i$ ,  $F_i$  le nombre de femmes dans ce même poste, et  $l$  le nombre de catégories de postes. L'indice  $R$  mesure l'écart de la distribution des postes par rapport à une représentation proportionnelle des hommes et des femmes dans les différentes catégories :

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

Dans un marché de travail parfaitement intégré,  $R$  est égal à 0.

*Source* : Calculs des auteurs à partir d'ILOSTAT, qui répartit les postes en sept catégories.

## Références bibliographiques

Blinder A.S. (1973) : « Wage Discrimination: Reduced Form and Structural Estimates », *The Journal of Human Resources*, vol. 8, n° 4, pp. 436-455.

Charles M. et Grusky D. (1995) : « Models for Describing the Underlying Structure of Sex Segregation », *American Journal of Sociology*, vol. 100, n° 4, pp. 931-71

Heckman J. (1979) : « Sample Selection Bias as a Specification Error », *Econometrica*, vol. 47, n° 1, pp. 153-161.

Meurs D. et S. Ponthieux (2006) : « L'écart de salaire entre les hommes et les femmes peut-il encore baisser », *Économie et Statistique*, n° 398-399, pp. 99-130.

Mulligan C.B. et Y. Rubinstein (2008) : « Selection, Investment, and Women's Relative Wages Over Time », *Quarterly Journal of Economics*, vol. 123, n° 3, pp. 1061-1110.

Oaxaca R. et M. Ransom (1999) : « Identification in Detailed Wage Decompositions », *The Review of Economics and Statistics*, vol. 81, n° 1, pp. 154-157.

Oaxaca R.L. (1973) : « Male-Female Wage Differential in the Urban Labor Markets », *International Economic Review*, vol. 14, n° 3, pp. 693-709.