



# TENDANCES ECONOMIQUES



Analyses et prévisions conjoncturelles



# Les salaires wallons sont-ils trop élevés ?

## Une étude économétrique des différences de salaires entre les régions

DOCQUIER Frédéric et LAURENT Sébastien

### Introduction

Les statistiques sociales belges par région nous indiquent que les salaires des ouvriers sont, en moyenne, plus élevés en Wallonie que dans les autres régions du pays. D'autre part, il est également clair que le taux de chômage wallon est plus élevé que celui de la Flandre. Dès lors, une conclusion hâtive pourrait amener d'aucuns à affirmer que l'excès d'offre de travail wallon est dû à un niveau trop élevé des salaires.

En utilisant les données du panel sur les ménages belges<sup>1</sup>, on étudie les disparités régionales des salaires en Belgique et on réfute l'idée selon laquelle les salaires wallons sont anormalement élevés.

Pour ce faire, on modélise une fonction de salaire dans laquelle l'appartenance à une région est une variable explicative potentielle. On procède alors à l'estimation économétrique des coefficients.

Afin d'éviter un biais d'estimation lié au fait que certains individus de notre échantillon ont un salaire nul (les non-travailleurs), on utilise la méthode d'estimation en deux étapes proposée par Heckman<sup>2</sup>. Cette méthode consiste à modéliser, en première étape, la probabilité d'avoir un travail (et donc, un salaire) en fonction des caractéristiques individuelles telles que l'âge, le niveau d'éducation, la situation familiale, etc ... . Ensuite, les résultats de cette

première étape sont utilisés dans la fonction de salaire, ce qui permet une estimation sans biais des coefficients interprétables.

D'une manière générale, on peut résumer nos conclusions de la manière suivante :

- premièrement, la probabilité de travailler est significativement plus élevée en Flandre que dans les autres régions;

- deuxièmement, l'appartenance à une région ne modifie pas significativement le niveau du salaire offert lorsqu'on prend en considération d'autres variables explicatives telles que l'ancienneté, le niveau d'éducation, le statut ou encore l'état civil. En conséquence, il n'y a pas de disparités significatives de politique de salaire entre les régions du pays;

- enfin, la méthode en deux étapes de Heckman se justifie pleinement dans la mesure où on montre que, statistiquement, l'absence de prise en compte des non-travailleurs aurait engendré un biais significatif sur les estimateurs.

La suite de cette étude s'organise comme suit: la section 2 présente les statistiques primaires du marché du travail par région. La section 3 présente les fondements théoriques de l'équation de salaire ainsi que la méthode d'estimation. La section 4 décrit l'échantillon et les résultats. Enfin, la section 5 conclut.

---

1 Voir Université de Liège (1992), « Panel study on Belgian Households », Point d'appui/Panel/Démographie familiale, printemps 1992.

2 Voir Heckman J. (1979), « Sample selection bias as a specification error », *Econometrica*, Vol 47, pp. 153-161.

## Statistiques primaires par région en Belgique

Les statistiques primaires<sup>3</sup> nous livrent une série de données désagrégées par région. On dispose de données relatives aux salaires offerts dans l'industrie ainsi que de nombreux indicateurs d'offre et de demande sur le marché du travail. Les tableaux 1 à 3 regroupent quelques-unes de ces données.

Comme l'illustre le tableau 1, le salaire brut moyen de l'ouvrier wallon est respectivement 6% et 9 % plus élevé que celui de son homologue bruxellois et flamand. Toutefois, cette différence est quelque peu atténuée par le fait que le salaire mensuel brut moyen des employés est légèrement plus faible en Wallonie.

Cette conclusion simpliste ne tient cependant pas compte des spécificités régionales du marché du travail. Ainsi, cette étude montre que les disparités salariales s'expliquent complètement par des structures différentes

Les tableaux 2 et 3 concernent la structure du marché<sup>4</sup>. Ils permettent de remarquer des différences régionales importantes dans les taux d'activité d'une part, dans les taux de chômage d'autre part. On constate que les taux d'activité sont nettement plus faibles à Bruxelles et, dans une moindre mesure, en Wallonie. Les taux de chômage quant à eux, traduisent l'effet inverse.

D'un point de vue strictement macroéconomique, la théorie néoclassique du marché du travail peut suggérer l'idée que l'importance des niveaux de salaires wallons est la cause première expliquant un tel écart de chômage entre la Wallonie et la Flandre. de marché du travail et les spécificités d'offre régionales. que les disparités salariales s'expliquent complètement par des structures différentes de marché du travail et les spécificités d'offre régionales.

**Tableau 1 - Salaires bruts dans l'industrie (en francs belges)**

	Région wallonne	Région flamande	Région bruxelloise
Salaire horaire brut des ouvriers dans l'industrie	435,52	399,34	411,3
Salaire mensuel brut des employés dans l'industrie	100.044	101.637	100.313

**Tableau 2 : Taux d'activité de la population masculine par classe d'âges**

	25-29	30-34	35-39	40-44	45-49
Royaume	89,7	92,4	92,3	91,8	90,1
Bruxelles	73,1	78,4	79,6	80,9	80,8
Flandre	93,3	95,2	94,8	94,0	92,3
Wallonie	88,2	91,2	91,3	90,9	88,3

**Tableau 3 : Taux de chômage des hommes par classe d'âges**

	25-29	30-34	35-39	40-44	45-49
Royaume	7,3	5,6	5,3	5,6	6,1
Bruxelles	13,6	10,9	10,7	9,9	11,4
Flandre	4,3	3,4	3,3	3,8	4,5
Wallonie	11,7	8,7	7,7	7,8	8,0

3 Les données ayant permis l'élaboration des tableaux et du graphique proviennent de : INS, « Communiqués hebdomadaires », INS, « Statistiques sociales » et INS, « Recensement de la population et des logements » 1991.

4 Taux activité = (Population active occupée + Demandeurs d'emploi) x 100 / Population.

Taux de chômage = Demandeurs d'emploi / Population active x 100.

## L'équation de salaire et la méthode d'estimation

Pour estimer les différences salariales entre les régions, il convient de répertorier l'ensemble des facteurs susceptibles d'exercer un effet sur le salaire. D'un point de vue strictement théorique, la théorie de l'économie de l'emploi nous livre un certain nombre de variables explicatives telles que le niveau d'éducation, l'ancienneté, le statut professionnel, le sexe, le pays d'origine...

Notre méthode d'investigation consiste à ajouter la région dans laquelle l'individu est employé comme variable explicative potentielle. Un coefficient significativement différent de zéro pour cette variable impliquerait une politique de salaire différente par région.

Dans cette section, on présente les éléments théoriques sous-jacents à la construction de notre équation de salaire et on discute une procédure d'estimation en deux étapes qui fournit des estimateurs sans biais.

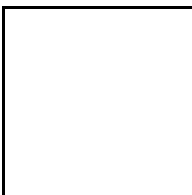
### Les variables explicatives potentielles

Dans son article empirique de 1974, Mincer propose une fonction de gains simple dont la spécification est devenue la référence pour les travaux économétriques<sup>5</sup>. Cette fonction exprime le logarithme du salaire,  $\log Y$ ; comme une fonction supposée croissante du niveau d'éducation de l'individu  $S_i$ ; de l'expérience professionnelle,  $E_i$ , et comme une fonction décroissante du carré de cette dernière. Pour l'individu  $i$ , on écrit donc

$$\log Y_i = \alpha + \beta_1 S_i + \beta_2 E_i + \beta_3 E_i^2 + \varepsilon_i \quad (1)$$

---

5 Voir Mincer J. (1974), « Schooling, experience and earnings », NBER, *Human behaviour and social institutions series*, 2, New York.

où  est le terme d'erreur aléatoire.

Cette fonction constitue notre point de départ pour l'étude des disparités régionales. D'emblée, précisons-en les fondements théoriques.

La *théorie du capital humain* justifie la présence du niveau d'éducation de l'individu,  $S$ , comme variable explicative. Selon cette théorie, les individus investissent en capital humain, ou en éducation, pour accroître leur position financière.

Cet investissement engendre deux effets opposés sur le revenu de cycle de vie de l'agent. D'une part, il implique un coût d'opportunité qui se traduit par un manque à gagner immédiat : si l'individu se forme, il le fait au détriment de son travail et, partant, des revenus potentiels immédiats<sup>6</sup>. D'autre part, l'individu qui se forme escompte un accroissement de sa productivité future et peut espérer obtenir de la sorte un emploi mieux rémunéré. Comme le montrent généralement les modèles économiques théoriques, l'individu a intérêt à investir une proportion de son temps telle que le gain marginal de revenu futur attendu soit juste égal à la perte marginale liée au coût de l'éducation.

En corollaire, la condition nécessaire pour que l'investissement soit rentable est qu'il existe une relation entre le niveau de capital humain et le revenu, cette relation mesurant en réalité le « taux de rendement de l'éducation ». Même si plusieurs études ont montré que ces taux de rendement privés ont chuté dans les pays développés durant les années 70, il apparaît

---

6 Notons que ce coût d'opportunité peut s'accompagner d'un coût monétaire ou financier dans la mesure où la formation peut nécessiter des frais considérables.

7 Voir, par exemple, Psacharopoulos G. (1981), « Returns to education: an updated international comparison », *Comparative education*, Vol 17, pp. 321-341, ou Pissarides C.A. (1982), « From school to university: the

clairement qu'il existe une relation positive forte entre le salaire et le niveau d'éducation. Dans notre équation de salaire, on retiendra onze catégories de diplômes allant du « sans diplôme » jusqu'aux « postuniversitaires ».

L'effet de séniorité explique la présence de l'expérience ou de l'ancienneté comme seconde variable explicative potentielle<sup>8</sup>. Il est empiriquement reconnu que les salaires augmentent avec l'âge ou, plus généralement, avec l'ancienneté, du moins jusqu'à un certain âge. Généralement, cela se traduit par un profil « en dos d'âne » de la fonction de salaire<sup>9</sup>.

Pendant longtemps, un grand nombre d'études empiriques basées sur l'idée que le salaire reflète la productivité individuelle, ont examiné le lien entre l'âge et la productivité. Cette voie d'explication n'aboutit pas à des résultats concluants et est fortement contestée aujourd'hui, plus particulièrement suite à l'émergence de la théorie des contrats implicites.

Cette nouvelle approche se fonde sur le principe qu'il est intéressant, pour la firme, de lier le salaire à l'ancienneté. En effet, le travailleur bénéficie d'une formation générale ou d'une formation spécifique au moment où il débute son activité professionnelle. Généralement, on peut supposer que sa productivité est plus basse pendant cette période et que cela entraîne des coûts de formation pour la firme. Pour éviter que le travailleur ne soit tenté d'exercer ses fonctions ailleurs une fois formé, il est rationnel que la firme le paie en dessous de sa productivité en début de carrière, puis au-dessus de sa productivité en fin de carrière.

Comment peut-on alors justifier un effet de concavité qui se traduit par un profil en dos d'âne. Lazear propose une explication

---

demand for post compulsory education in Britain », *Economic Journal*, Vol 92, pp. 654-667.

<sup>8</sup> En approximation, on mesure l'expérience comme la différence entre l'âge et l'âge de fin d'étude.

<sup>9</sup> Par profil en dos-d'âne, concept largement discuté par Lazear (1979) et Mincer (1974), il faut entendre un effet positif de l'expérience sur le salaire dans un premier temps et une diminution de ce dernier en fin de cycle de vie du travailleur. En pratique, nous pouvons traduire cet effet par le caractère concave de la fonction de salaire estimée.

convaincante<sup>10</sup>. Si le travailleur est payé au-dessus de sa productivité en fin de carrière, il va de soi que la firme a intérêt à ce qu'il y ait une date de fin au contrat de travail. Autrement dit, au niveau individuel, il est intéressant pour la firme de diminuer le différentiel salaire/productivité à l'approche de l'âge de la retraite. Le travailleur peut choisir de rester dans l'entreprise, mais à un salaire moindre. Les estimations empiriques de Mincer montrent que le salaire maximal est, en moyenne, obtenu après 34 ans d'expérience, soit à un âge proche de 55 ans.

A ces deux variables explicatives, éducation et ancienneté, on peut encore ajouter la nature de l'emploi considéré. Ainsi, la manière dont les indépendants sont rémunérés diffère de celle dont les salariés le sont. Il est également probable que les variations conjoncturelles et structurelles impliquent que les conditions sur le marché du travail varient fortement entre les secteurs industriels, ce qui peut expliquer des différences de salaires selon la branche d'activité. Dans cette étude, on ne dispose pas de données relatives au secteur d'activité. De plus, on choisit de travailler uniquement sur les salariés et on exclut les indépendants. Toutefois, on tient compte du fait que le travailleur exerce sa profession dans le secteur privé ou dans le secteur public. On sait, en effet, que le secteur public se caractérise généralement pas une plus grande sécurité d'emploi mais des salaires plus bas. On teste cette hypothèse.

Enfin, on se limite ici aux salariés de sexe masculin mais on modélise la situation familiale de l'agent au moyen de deux variables : l'état civil et le lien avec le chef du ménage. Ces variables influencent à tout le moins la décision de travailler de l'individu.

## L'équation de salaire

En fonction des variables explicatives potentielles, notre équation de salaire est la suivante :

$$\boxed{\hspace{10em}} \quad (2)$$

---

<sup>10</sup> Voir Lazear E.P. (1979), « Why is mandatory retirement », *Journal of Political Economy*, 87, pp.1261-1284.

où  $w_i^*$  est le salaire de l'individu  $i$ ,  
 $E_i$  est une variable binaire  
 spécifiant si l'individu  $i$  appartient ou non à la  
 classe de diplôme  $d$ ,  $E_i$  représente son  
 ancienneté mesurée par la différence entre l'âge  
 actuel et l'âge de fin de scolarité,  $PUB_i$  est  
 une variable binaire spécifiant si l'individu  
 travaille dans le secteur public ou non et  
 $REG_{i,r}$  est une variable binaire spécifiant si  
 l'individu travaille dans la région  $r$  (région  
 Bruxelles-capitale, si  $r=B$ , ou région flamande,  
 si  $r=F$ )

## Le biais de sélection et la méthode en deux étapes

Lorsqu'il cherche à estimer une fonction de salaire telle que l'équation (2), l'économètre est confronté au problème technique suivant : certains individus de l'échantillon sont sans travail et ont donc un salaire nul. S'ils sont chômeurs, invalides, pensionnés ..., ces individus bénéficient de revenus de remplacement. Sinon, ils vivent de leurs revenus financiers ou d'autres ressources possibles. Qu'ils soient demandeurs d'emploi ou non, ces individus sans salaire doivent être traités de manière appropriée dans notre estimation. Peut-on utiliser la technique habituelle des moindres carrés en leur présence? Faut-il, au contraire, les éliminer de notre échantillon?

Les développements économétriques d'Heckman montrent qu'aucune de ces deux options n'est satisfaisante. Précisons donc la démarche correcte en faisant appel à la théorie économique de base.

Si on veut caractériser la politique de salaire, on doit expliquer le salaire offert par les entreprises,  $w_i^*$ , en fonction des explicatives. Du point de vue économique, l'individu  $i$  choisit de travailler si le salaire qui lui est offert,  $w_i^*$ ,

est supérieur à son salaire de réserve,  $s_i^*$ . Or, ces deux variables ne sont pas observables<sup>11</sup>. Le problème qui se pose ici est que notre échantillon de salaires observés,  $w_i$ , n'est pas représentatif de la distribution complète des salaires offerts. On ne dispose de données que pour les agents dont le salaire offert est supérieur au salaire de réserve. Pour les autres agents, le salaire observé est nul alors que le salaire offert peut être, quant à lui, différent de zéro.

En d'autres termes, la variable observable de salaire vaut

$$w_i = \begin{cases} w_i^* & \text{si } w_i^* > s_i^* \\ 0 & \text{sinon} \end{cases} \quad (3)$$

On peut montrer que l'application de la méthode des moindres carrés ordinaires sur l'ensemble de l'échantillon (y compris les salaires nuls) débouche sur des estimateurs biaisés. De même, si on élimine les sans-travail, l'estimateur est également biaisé. Toutefois, dans ce cas, le biais dépend étroitement de la probabilité, pour l'individu  $i$ , d'avoir un salaire positif. On note cette probabilité

$$P(Trav = 1) = P(w_i^* > s_i^*) \quad (4)$$

Pour remédier à cette situation, Heckman propose une méthode en deux étapes qui fournit un estimateur sans biais des coefficients. Cette méthode consiste à estimer le facteur de biais dans une première étape, ce qui servira à estimer l'équation de salaire dans une seconde étape. Plus formellement, cette méthode consiste à séparer les parties qualitative (probabilité de travailler) et quantitative (équation de salaire) du modèle.

## Mise en pratique

Pour évaluer la probabilité de travailler, on construit un modèle PROBIT. Ceci revient à modéliser la probabilité que le salaire offert soit supérieur au salaire de réserve en fonction des caractéristiques individuelles telles que le niveau de diplôme, le nombre d'enfants, l'état civil, l'âge, ... On écrit donc

$$P(w_i^* > s_i^*) = \Phi(X_i \beta) \quad (5)$$

où  $\Phi$  est la fonction de distribution de la loi normale centrée réduite,  $\beta$  est un vecteur de paramètres à estimer et  $X_i$  est un vecteur de

<sup>11</sup> L'étoile traduit la non-observabilité de la variable.

variables explicatives de la probabilité de travailler.

En seconde étape, on sélectionne uniquement les travailleurs avec salaire positif et on estime, par moindres carrés ordinaires, l'équation de salaire (2) en y ajoutant une variable supplémentaire liée au modèle probit de la première étape:

$$MIL_i = \frac{\sigma_i \varepsilon_i}{\sigma_i} \quad (6)$$

où  $\sigma_i$  est l'écart type des salaires observés,  $\varepsilon_i$  est la densité de la loi normale et  $\hat{\sigma}_i$  est le paramètre estimé lors de la première étape. Cette variable est généralement appelée « ratio de Mill » et corrige le biais de sélection.

## Résultats de l'estimation de la fonction de gains de Mincer

Notre estimation porte sur les données issues de l'enquête « Point d'Appui, Panel, Démographie Familiale » réalisée au Printemps 1992. Suite à cette enquête de grande envergure, nous disposons de plus de 8000 observations regroupées sous le nom de « Panel Study on Belgian Households » ou en abrégé, PSBH.

### Sélection de l'échantillon.

La sélection de notre échantillon nous permettra de travailler avec un sous-ensemble de 1594 observations. L'échantillon retenu comprend les hommes âgés de 25 à 50 ans appartenant aux catégories socioprofessionnelles suivantes : chômeurs (indemnisés ou non), ouvriers (qualifiés ou non), employés, cadres et autres salariés. Nous excluons donc de notre échantillon les pensionnés, étudiants, indépendants, agriculteurs ou encore les professions libérales.

### Quelques statistiques

L'étude du salaire mensuel nous pousse à accepter qu'il existe bel et bien un différentiel régional en terme de moyennes salariales. La région bruxelloise est à ce sujet caractérisée par un salaire moyen relativement élevé (BEF 56395). Viennent ensuite la Wallonie avec BEF 53162 et la Flandre avec BEF 51974.

Le tableau 4 présente succinctement la composition de l'échantillon sélectionné en distinguant le Royaume des trois régions administratives. Que peut-on en inférer ?

#### 1. Concernant les classes d'âge

La prise en considération de la répartition en classes d'âge permet de constater que l'échantillon est composé majoritairement d'individus de 25 à 34 ans. Cependant, il est

crucial de remarquer les disparités régionales dans la distribution de la population par classes d'âge. La Wallonie est à ce point caractérisée par une population active plus âgée que Bruxelles et que la Flandre.

#### 2. Concernant les catégories de diplôme

Une brève étude de la répartition des catégories de diplôme nous montre que la région bruxelloise est nettement plus scolarisée que les régions wallonne et flamande. C'est en effet à Bruxelles que l'on trouve le pourcentage le plus élevé d'individus disposant d'un diplôme de type universitaire<sup>12</sup>.

3. *Catégorie socioprofessionnelle* : La composition de notre échantillon reflète exactement les spécificités régionales en terme d'emploi. Une comparaison avec les statistiques sociales régionales primaires nous permet de penser que notre échantillon est assez représentatif. La distribution (de la population étudiée) par catégorie socioprofessionnelle nous confirme que le taux de chômage flamand (4,05 %) est nettement inférieur aux taux rencontrés en Wallonie (8,95 %) et en région bruxelloise (9,27 %). En outre, une étude plus approfondie de ces données nous révèle que Bruxelles comporte significativement plus de cadres que les deux autres régions mais beaucoup moins de travailleurs manuels, ce qui semble très plausible.

4. *Statut* : La distinction effectuée entre secteur privé et public illustre des différences régionales particulièrement intéressantes, à savoir la prédominance marquée de la Flandre pour le

---

<sup>12</sup> Les pourcentages de diplôme de candidature, licence et postuniversitaire valent respectivement 22,9 % à Bruxelles, contre 9,6 % en Flandre et 11,9 % en Wallonie.

secteur privé (75,32 % contre 24,68 % pour le public) alors que la Wallonie révèle une attirance nettement moindre pour ce secteur (59,01 % contre 40,98 % pour le secteur public).



**Tableau 4 : Présentation de l'échantillon (en pourcentages)**

		<b>Belgique</b>	<b>Bruxelles</b>	<b>Flandre</b>	<b>Wallonie</b>
<b>Observations</b>		100	12,86	46,49	40,65
<b>Etat civil</b>					
CIV1	Couple	78,9	68,47	82,88	77,6
CIV2	Célibataire	11,8	19,21	9,51	11,98
CIV3	Veuf	0,4	0,99	0,27	0,47
CIV4	Divorcé	6,4	7,88	5,16	7,47
CIV5	Séparé	2,5	3,45	2,17	2,49
<b>Nationalité</b>					
	Belgique	90,4	69,8	97,57	88,49
	Europe	6,7	17,33	1,48	9,33
	Asie	1	1,98	0,54	1,24
	Afrique	1,6	9,9	0,13	0,78
	Amérique	0,1	0,5	0	0,16
	Autre	0,2	0,5	0,27	0
<b>Classe d'âge</b>					
	25-29 ans	19,5	21,95	20,78	17,28
	30-34 ans	26,2	23,41	28,07	24,85
	35-39 ans	18,3	18,54	16,73	19,91
	40-44 ans	18,8	18,54	17,27	20,52
	45-50 ans	17,3	17,56	17,14	17,44
<b>Catégorie de diplôme</b>					
DIP1	Sans diplôme	1,8	2,93	1,48	1,7
DIP2	Primaire, 4ème degré, spécial	8,5	9,27	8,91	7,87
DIP3	Secondaire inférieur	20,8	14,63	21,46	22,07
DIP4	Secondaire supérieur	28,7	22,44	28,34	31,02
DIP5	Formation professionnelle + apprentissage + secrétaires + classes moyennes	5,7	4,39	6,75	4,94
DIP6	Supérieur de type court	14,3	15,12	15,11	13,12
DIP7	Supérieur de type long	5,1	5,37	5,67	4,48
DIP8	Universitaire candidature	1,2	3,41	0,94	0,77
DIP9	Universitaire licence	8,1	12,2	6,88	8,18
DIP10	Postuniversitaire	2,9	7,32	1,75	2,93
DIP11	Autres	2,8	2,93	2,7	2,93
<b>Catégorie socioprofessionnelle</b>					
	Chômeur (indemnisé ou non)	6,7	9,27	4,05	8,95
	Ouvrier non qualifié	9,4	8,29	12,69	6,02
	Ouvrier qualifié	29	18,54	31,58	29,48
	Employé	29,2	27,8	28,88	29,94
	Cadre	20,8	28,29	18,89	20,68
	Autre salarié	4,8	7,8	3,91	4,94
<b>Statut</b>					
PRIV	Privé	67,8	67,03	75,32	59,01
PUB	Public (sous contrat ou nommé)	32,2	32,97	24,68	40,98
<b>Lien avec le chef de ménage</b>					
LIEN0	Chef de ménage	88,3	83,9	89,47	88,43
LIEN1	Partenaire	3,5	8,29	2,56	3,09
LIEN2	Enfant, frère, sœur, + autres	8,2	7,81	7,95	8,32

C'est sur la base de cet échantillon de 1594 individus qu'on va estimer l'équation de salaire. Les sections suivantes présentent les résultats des deux étapes de la méthode d'Heckman.

## Estimation de la fonction probit

La première étape consiste à estimer la probabilité de travailler au moyen d'une fonction probit dont la modélisation est la suivante :

$$P(Tr_i = 1) = \Phi(\beta_0 + \beta_1 DIP_i + \beta_2 AGE_i + \beta_3 CIV_i + \beta_4 LIEN_i + \beta_5 SALC_i + \beta_6 ENF_i + \beta_7 REG_i) \quad (7)$$

où *DIP*, *CIV* et, *LIEN* sont des vecteurs de variables auxiliaires caractérisant la catégorie de diplôme, l'état civil et le lien avec le chef de ménage (dont les subdivisions sont définies au tableau 1 de la section 1.2 ), *SALC*, et *REG* sont également des vecteurs de variables auxiliaires correspondant à la classe de salaire du conjoint (ou du partenaire) 13 et à la région<sup>14</sup> (lieu d'habitation et non le lieu de travail), *ENF* est une variable déterminant le nombre d'enfants de l'individu et *AGE*, l'âge de l'individu.

Il est à noter que la constante caractérise un individu ne disposant pas de diplôme (*DIP0*), vivant en couple (*CIV1*), chef de ménage (*LIEN0*), dont le conjoint ne dispose pas de salaire et habitant en Wallonie.

L'application de cette fonction à l'échantillon<sup>15</sup> nous livre les résultats présentés dans le tableau 5.

---

13 Les variables auxiliaires composant le vecteur *SALC* sont :

*SALC0* si le conjoint gagne moins de 30000 FB/mois (dans la constante)

*SALC1* si le conjoint gagne entre 30000 et 60000 FB/mois

*SALC2* si le conjoint gagne + de 60000 FB/mois.

14 Les variables auxiliaires composant le vecteur *REG* sont :

*REGW* si l'individu habite en Wallonie (dans la constante)

*REGB* si l'individu habite à Bruxelles

*REGV* si l'individu habite en Flandre.

15 Afin d'assurer la représentativité de l'échantillon, il nous est apparu opportun de pondérer les données individuelles en fonction du sexe, de l'âge et de la région en prenant comme référence les individus de l'Enquête

La première remarque intéressante est que la possession d'un diplôme universitaire accroît significativement la probabilité de travailler. Toutefois, l'obtention d'un diplôme de candidature n'accroît pas la probabilité de travailler par rapport au diplôme de l'enseignement secondaire.

On constate également que l'importance du salaire du conjoint accroît la probabilité de travailler. Ceci peut paraître anormal dans la mesure où la théorie néoclassique du marché du travail indique que le salaire du partenaire masculin accroît le salaire de réserve de la femme (ce salaire ayant un effet quasiment 16 similaire à l'accroissement du revenu non salarial de la femme) et donc diminue sa probabilité de travailler.

Ici, nous pouvons expliquer cette corrélation positive par le fait que l'homme travaille quoi qu'il arrive, que les couples sont généralement formés d'individus dont les qualifications sont de même niveau et donc, que le salaire de la femme peut être en quelque sorte un indicateur du niveau d'éducation de l'homme.

Enfin, nous pouvons remarquer un effet régional très significatif en faveur de la Flandre. Ainsi, le seul fait de résider en Flandre accroît de 4 % la probabilité de travailler. Cet effet régional traduit un ensemble d'éléments très différents dont les principaux sont peut-être :

- les disparités éventuelles de fonctionnement des marchés ;
- des éléments propres à la demande de travail (compétitivité régionale, importance et développement des secteurs d'activité, poids des différents secteurs industriels, etc...).

---

des Forces de Travail (INS, Eurostat). Des problèmes de sur- ou de sous-représentation des trois régions du pays ainsi que des catégories socioprofessionnelles nous ont donc poussés à intégrer une variable de pondération dans toutes nos régressions. Pour plus de précisions, voir « Proposition d'une procédure de pondération de la première vague du Panel Démographie Familiale », Foidart-Orban-Perelman (disponible au CREPP)

16 Nous parlons d'effet quasiment identique et non identique car l'effet de l'accroissement du revenu non salarial et le salaire du conjoint diffèrent essentiellement par l'existence de coûts fixes supplémentaires dans le cas d'un ménage.

**Tableau 5 - Estimation du modèle probit (1ère étape)**

<u>Paramètre</u>	<u>Estimateur</u>	<u>t-statistique</u>		<u>P-value</u>
C	1.57135	3.19559	**	[.001]
DIP2	-.552494	-1.61228		[.107]
DIP3	.062378	.186111		[.852]
DIP4	.264019	.786866		[.431]
DIP5	.150640	.389842		[.697]
DIP6	.526335	1.44828		[.148]
DIP7	.479605	1.13673		[.256]
DIP8	.280848	.456908		[.648]
DIP9	.957183	2.17228	*	[.030]
DIP10	.889086	1.60299		[.109]
DIP11	.727467	1.35318		[.176]
I_AGES	-.957150E-02	-1.17273		[.241]
CIV2	-.379792	-1.91074		[.056]
CIV3	-.961113	-1.79250		[.073]
CIV4	.081646	.360582		[.718]
CIV5	-.184298	-.613414		[.540]
LIEN1	-.386662	-1.35754		[.175]
LIEN2	-.369624	-1.82524		[.068]
SALC1	.484954	3.55941	**	[.000]
SALC2	.589428	2.36938	*	[.018]
M_ENFA	-.095587	-1.82656		[.068]
REGV	.390262	3.26030	**	[.001]
REGB	-.016762	-.105498		[.916]
LOG OF LIKELIHOOD FUNCTION	=	-330.225		
NUMBER OF OBSERVATIONS	=	1594		
PERCENT CORRECT PREDICTIONS	=	.932873		

Les autres variables disponibles dans la base de données se sont avérées fortement non significatives<sup>17</sup>. Ainsi, l'état civil, le lien avec le chef de ménage, le nombre d'enfants (variable très importante pour l'estimation de la probabilité de travailler des femmes) et la nationalité (non conservée dans l'estimation de notre dernière fonction probit) n'ont pas d'influence significative sur la probabilité de travailler.

l'équation de salaire (2) dans laquelle, rappelons-le, on ajoute le ratio de Mill comme variable explicative potentielle :

## Estimation de la fonction de salaire

La seconde étape consiste à éliminer les individus dont le salaire est nul (et donc conserver seulement les salariés) estimer

---

<sup>17</sup> Nous avons opté pour une valeur ? de 5%. Il est bien entendu que certains paramètres sont significatifs pour une valeur de ? légèrement supérieure.

$$\ln Y_i = a_0 + a_1 DIP + a_2 E_i + a_3 E_i^2 + a_4 PUB_i + a_5 REG + a_6 MIL + \epsilon_i \quad (2')$$

Par convention, nous avons décidé de considérer comme constant, le logarithme du salaire de l'individu non pourvu de diplôme (DIP1), travaillant dans le secteur privé (PRIV) et habitant en région wallonne (REGW).

Le tableau suivant nous livre les résultats du test effectué sur les 1487 salariés<sup>18</sup>.

---

<sup>18</sup> Le nombre de « sans-travail » valant 107.

**Tableau 6 : Estimation de la fonction de salaire (2e étape)**

Variable	Coefficient	Erreur	t-statistique	P-value
C	10.3151	.075503	136.617	** [.000]
DIP2	-.036425	.068846	-.529083	[.603]
DIP3	.028490	.065412	.435537	[.667]
DIP4	.159382	.066015	2.41434	* [.015]
DIP5	.027063	.070317	.384878	[.701]
DIP6	.320899	.067231	4.77306	** [.000]
DIP7	.450240	.071979	6.25518	** [.000]
DIP8	.183287	.148169	1.23701	[.214]
DIP9	.569091	.070370	8.08718	** [.000]
DIP10	.606613	.078642	7.71358	** [.000]
DIP11	.075999	.092076	.825396	[.414]
EXPL	.029003	.435825E-02	6.65479	** [.000]
EXPSQ	-.419282E-03	.117156E-03	-3.57884	** [.000]
PUB	-.034034	.014907	-2.28311	* [.021]
REGV	-.010496	.015856	-.661970	[.516]
REGB	.011274	.026193	.430411	[.670]
@MILLS	-.184500	.080667	-2.28718	* [.021]

Number of observations: 1487

Log of likelihood function = -155.826

Schwarz Bayes. Info. Crit. = -2.58907

R-carré ajusté = .345042

Standard Errors are heteroskedastic-consistent (HCTYPE=2).

Première constatation, le  $R^2$  est proche de 35%, ce qui constitue un excellent résultat dans un tel type d'analyse si l'on pense à l'importante hétérogénéité qui existe sur le marché du travail et si l'on se réfère aux résultats obtenus par Mincer (le  $R^2$  obtenu par Mincer étant de 30 %). Plusieurs coefficients sont significatifs<sup>19</sup>. Premièrement, le niveau de diplôme influence généralement positivement le salaire offert. Il est toutefois intéressant de remarquer la faiblesse de l'accroissement du salaire attendu pour un étudiant ayant terminé ses études supérieures et qui décide de travailler en faisant valoriser un diplôme de candidature universitaire.

Nous pouvons conforter nos anticipations quant aux signes attendus de l'expérience et de son carré. Ainsi, l'expérience a bel et bien un effet positif sur le salaire offert. De plus, le profil en « dos-d'âne » est également vérifié étant donné le signe négatif et significatif du coefficient relatif au carré de l'expérience. Le corollaire de ce résultat est que le salaire maximum est atteint

après 34 années d'expérience, soit à un âge proche de 55 ans<sup>20</sup>.

Ensuite, nous remarquons que le statut public entraîne une baisse de salaire significative. C'est là encore une fois la confirmation empirique d'un effet attendu dans la mesure où il est généralement reconnu que le secteur public assure un travail plus stable, au prix d'une perte de rémunération par rapport au secteur privé.

La quatrième remarque doit attirer tout particulièrement notre attention, car elle est l'objet principal de notre travail. Nous avons mis à jour l'existence d'un différentiel de salaire moyen régional. Cependant, au vu des résultats obtenus dans l'estimation de notre fonction de salaire, on constate qu'il n'y a pas d'effet régional significatif dans la politique de salaire offert<sup>21</sup>.

---

20 Hypothèse faite qu'en moyenne, un jeune obtient son premier emploi aux alentours de 20 ans.

21 Cette conclusion peut être confortée par le fait que ce vecteur régional n'a jamais été significatif tout au long de nos très nombreuses estimations (la spécification du

---

19 Pour un seuil de significativité de 5 %.

En résumé, bien que l'appartenance à une région particulière puisse influencer la probabilité de travailler, il est faux de croire qu'il existe des différences de salaire non fondées entre les régions en Belgique. Les disparités statistiques développées au tableau 4 expliquées par des différences en terme de niveau de diplôme, des structures de marché différentes mais aussi par des structures par âge de la population active différentes.

Enfin, l'inverse du ratio de Mill exerce effectivement un effet significatif, ce qui implique que sa non-prise en compte lors de la modélisation du modèle aurait conduit à des biais significatifs dans les estimateurs.

## Conclusions

Les deux équations estimées sur base du panel des ménages nous enseignent que l'appartenance à une région peut effectivement influencer la probabilité d'obtenir un emploi rémunéré mais qu'elle n'influence pas significativement le niveau de salaire offert. Les disparités régionales constatées sur les niveaux moyens de revenu des salariés trouvent donc explication dans d'autres variables qui, quant à elles, ont des distributions statistiques différentes selon les régions.

La première de ces variables, et qui à nos yeux est la plus importante, est le niveau d'éducation. Si l'on se réfère au tableau 1 de ce chapitre, on remarque que Bruxelles est caractérisée par une population nettement plus qualifiée que la Wallonie et surtout que la Flandre. Comme nous l'avons déjà fait remarquer, l'échantillon bruxellois est composé à 22,9 % d'universitaires contre 11,8 % et 9,6 % respectivement en Wallonie et en Flandre. Nous pouvons en déduire une part explicative relativement importante du niveau de diplôme dans le différentiel de salaire moyen en faveur de Bruxelles.

Le deuxième élément en importance permettant d'expliquer les différences régionales est certainement le niveau d'expérience et donc la structure par âge de la population belge. En réalité, si nous approfondissons quelque peu l'analyse statistique, nous pouvons mettre à jour

---

modèle présenté ici étant celle fournissant les meilleurs résultats).

un différentiel régional important en terme de moyenne d'âge. En effet, la Wallonie est caractérisée par la population la plus âgée (36,9 ans contre 36,4 ans à Bruxelles et 36,1 ans en Flandre). Si l'on se réfère aux théories de Lazear et Carmichael et donc à la progressivité du salaire avec l'ancienneté<sup>22</sup>, nous pouvons en partie expliquer le différentiel entre la Wallonie et la Flandre par cette différence structurelle. De plus, si l'on affine encore un peu l'étude statistique, on remarque que cette caractéristique régionale est conservée au niveau de l'expérience moyenne<sup>23</sup>, la Wallonie disposant de travailleurs plus « expérimentés ». Il est donc clair que les différences régionales en faveur de la Wallonie et en terme de salaires moyens peuvent être en partie expliquées par les différences régionales de structure par âge de la population.

Enfin, on a vu que l'appartenance au secteur public impliquait une baisse de salaire par rapport au secteur privé. D'après le Tableau 1, on voit que la part des fonctionnaires publics est nettement plus importante en Wallonie que dans les autres régions<sup>24</sup>. En conséquence, les écarts de distribution régionale des emplois du secteur public ont pour effet de diminuer les disparités salariales régionales.

---

22 Voir Lazear, op.cit. note 9, et Carmichael L. (1983), « Firm specific human capital and promotion ladders », *Bell Journal*, 14, pp. 251,258.

23 Les moyennes d'expérience sont respectivement de 16,25 années pour Bruxelles, 17.25 années pour la Flandre et 17.67 années pour la Wallonie.

24 Près de 40 % des salariés wallons étant employés dans le secteur public alors que ce pourcentage s'élève à 25 % en Flandre et à 33 % à Bruxelles.

## Annexe : Simulations illustratives

Pour illustrer le biais issu de la non-prise en compte des chômeurs dans l'estimation des salaires offerts, on développe un exemple avec données stochastiques simulées.

Ainsi, supposons que les individus ont un salaire de réservation qui dépend du nombre d'années d'études  $Etude_i$ , de leur âge  $Age_i$  et du nombre d'enfants  $Enf_i$  :

$$s_i^* = 90 + 2,5 \cdot Enf_i + 5,0 \cdot Etude_i + 0,3 \cdot Age_i + \eta_i$$

La fonction de salaire offert dépend de l'âge et du nombre d'années d'études :

$$w_i^* = 55 + 1,2 \cdot Age_i + 4,0 \cdot Etude_i + \varepsilon_i$$

Le problème est alors d'estimer la fonction de salaire lorsqu'on n'observe pas les variables  $w_i^*$

et  $s_i^*$  mais seulement le salaire effectif

$$w_i = w_i^* \text{ si } w_i^* > s_i^* \text{ et } w_i = 0 \text{ sinon.}$$

En simulation, les variables explicatives sont générées aléatoirement. L'échantillon simulé porte sur 2000 individus. Pour chacun d'eux, le nombre d'années d'études est une variable uniforme sur [0,6]. Le nombre d'enfant est la partie entière d'une variable aléatoire qui suit une loi de Poisson de moyenne 1. L'âge est distribué selon une loi normale de moyenne 35 et d'écart type 5. Enfin, les termes d'erreurs additifs sont distribués selon une loi normale bivariée de la forme

$$\begin{pmatrix} \eta_i \\ \varepsilon_i \end{pmatrix} \sim N \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 25 & 9 \\ 9 & 9 \end{pmatrix}$$

La première méthode d'estimation consiste à régresser le salaire observé sur les explicatives en prenant tout l'échantillon, y compris les salaires nuls. On obtient les résultats suivants.

### Estimation sur l'ensemble des données

Var	Coeff	Ecart-type	t-stat
C	-40.0976	4.23998	-9.45703
AGE	3.81674	.115239	33.1202
ETUDE	2.26465	.338045	6.69925

On voit évidemment que les coefficients estimés n'ont rien en commun avec les paramètres structurels du modèle vrai.

La seconde méthode consiste à opérer les moindres carrés ordinaires sur les salaires non-nuls et d'éliminer ainsi les non-travailleurs

(soient 172 observations sur 2000) de notre échantillon. On obtient ce qui suit.

### Estimation sur salaires positifs

Var	Coeff	Ecart-type	t-stat
C	58.5918	.966869	60.5995
AGE	1.10637	.026204	42.2216
ETUDE	4.07597	.068566	59.4461

On constate l'émergence d'un biais important dans les coefficients estimés. Ainsi, l'effet de l'âge est sous-estimé tandis que celui des années d'études est surestimé. La constante estimée quant à elle s'avère plus élevée que dans le modèle.

Enfin, si on applique la méthode d'Heckman, on construit un modèle PROBIT en première étape qui explique la probabilité de travailler, c'est-à-dire la probabilité que le salaire offert soit supérieur au salaire de réservation. On dégage le tableau suivant.

### Estimation du Probit (1ere étape)

Param	Coeff	Ecart-type	t-stat
C	-9.75481	.724912	-13.4565
AGE	.416383	.027418	15.1863
ETUDE	-.277626	.043383	-6.39938
ENF	-.673648	.074544	-9.03695

Cette première étape permet d'estimer la variable du deuxième terme de l'équation théorique de régression. La prise en compte de cette variable (notée MILL) dans l'estimation de la fonction de salaire corrige le biais de sélection. En deuxième étape, on dérive les estimateurs suivants.

### Estimation de la fonction de salaire (2e étape)

Var	Coeff	Ecart-type	t-stat
C	55.8844	1.14194	48.9379
AGE	1.17948	.030916	38.1508
ETUDE	4.03020	.069012	58.3984
MIL	3.18901	.724681	4.40057

On constate donc que la méthode en deux étapes fournit des estimateurs très proches des valeurs réelles des paramètres. En outre, la variable MILL est largement significative, ce qui montre bien que le biais de sélection jouait un rôle important dans l'estimation.