

**FACULTES UNIVERSITAIRES NOTRE-DAME DE LA PAIX LOUVAIN
SCHOOL OF MANAGEMENT**

EFASM005 - Séminaire d'économétrie et statistique appliquée au Mémoire

***D'où viennent les inégalités de salaires entre hommes
et femmes en Belgique?***



Année académique 2010-2011

Sous la direction de Jean-Yves GNABO

Assistante: Anne Michels

Remerciements

En préambule à cette partie quantitative du mémoire, nous souhaiterions adresser nos remerciements les plus sincères aux personnes qui, par leur enseignement, leur soutien et leurs précieux conseils nous ont aidés à sa réalisation.

Nous tenons tout d'abord à remercier sincèrement Monsieur Jean-Yves Gnabo, Professeur en Econométrie et Directeur du Séminaire d'économétrie et de statistiques appliquée au mémoire, pour son écoute, ses judicieux conseils afin que le travail soit réalisé dans un esprit scientifique rigoureux, sa disponibilité tout au long de la réalisation de ce travail malgré ses charges académiques et professionnelles.

Notre gratitude va à Madame Anne Michels, Assistante dans le cadre du Séminaire d'économétrie et de statistiques appliquée au mémoire, qui nous a guidés dans la démarche à suivre pour l'estimation des données. Nous la remercions pour sa gentillesse, sa patience, sa disponibilité, son écoute attentive et ses vastes connaissances en économétrie.

Nous désirons remercier Kea Tijdens, Sanne van Zijl, Melanie Hughie-Williams, Maarten van Klaveren, Stéphanie Steinmetz. Auteurs de l'enquête de 2010 dénommée *Codebook and explanatory note on the WageIndicator dataset, a worldwide, longitudinal, multilingual web and paper survey on work and wages* et travaillant au sein de l'University of Amsterdam et de l'Amsterdam Institute for Advanced Labour Studies, ils nous ont généreusement transmis la base de données constituant le point de départ du présent travail.

Nous témoignons notre reconnaissance à Madame Marie-Paule Urbain, Ministère Fédéral de l'Emploi et du Travail, Direction de l'Egalité des Chances, à Monsieur Pieter Vermeulen, Direction Générale des Statistiques du Gouvernement fédéral et à Monsieur Didier Dupré, Eurostat Unit F2, au Professeur Dimitri Mortelmans, Université d'Anvers, pour leurs conseils dans la recherche de notre base de données.

Enfin, nous exprimons notre reconnaissance envers nos familles respectives pour leur soutien, leur encouragement, leur compréhension tout au long de notre démarche.

Table des matières

Introduction	4
1. Revue de la littérature.....	6
2. Démarche Empirique	9
2.1. Base de données	9
2.1.1. <u>La base de données</u>	9
2.1.2. <u>Les caractéristiques des données</u>	10
2.1.3. <u>Lien entre la variable que l'on cherche à expliquer et les variables explicatives</u> ..	14
2.1.4. <u>Lien entre les variables explicatives</u>	14
2.2. Méthode utilisée.....	14
2.2.1. <u>La méthode des Moindres Carrés Ordinaires</u>	14
2.2.2. <u>La méthode d'Oaxaca-Blinder (1973)</u>	15
3. Compte rendu de l'analyse	17
3.1. Analyse descriptive	17
3.2. Analyse empirique	19
3.2.1. <u>Régression directe sur les deux groupes hommes-femmes</u>	19
3.2.2. <u>Régression sur le groupe hommes</u>	21
3.2.3. <u>Régression sur le groupe femmes</u>	23
3.2.4. <u>Méthode d'Oaxaca-Blinder</u>	24
Conclusion	26
Annexes	28
Bibliographie.....	40

Introduction

La féminisation de la population active est progressivement devenue une réalité depuis plusieurs décennies. Toutefois, il s'avère qu'à l'heure actuelle, la gente féminine perçoit une rémunération inférieure à celle de la gente masculine et que le sexe joue un rôle déterminant dans la position professionnelle. En 1973, Ronald L. Oaxaca et Alan S. Blinder, économistes américains, ont développé une méthode permettant de mesurer la différence de salaire imputable à la discrimination¹. Celle-ci est explicitée au point 2.2.2., page 15 du présent travail. C'est sur base de leur méthode et des développements de celle-ci que de nombreuses études économétriques ont été menées. Les résultats de ces études ont révélé que la discrimination envers les femmes est bien l'une des causes de la différence salariale entre le sexe faible et le sexe fort.

La discrimination salariale envers les femmes serait notamment liée à différents facteurs tels que *les choix politiques au cours du temps*. Au début du 19e siècle, avant la révolution industrielle, les métiers des femmes étaient très différents des métiers exercés par les hommes. Elles étaient "*fileuse, couturière, boutonnière, dentellière, fille de laiterie, femme de chambre, ...*"². Le salaire des hommes permettait de satisfaire les besoins du foyer (nourriture, logement, ...) tandis que celui des femmes constituaient plutôt un salaire d'appoint. De plus, le "Code Napoléon de 1804" faisait part de leur infériorité en les privant de certains droits. Au cours du 19e siècle, les économistes se sont penché sur le travail des femmes et à la fin du siècle, des lois sur les femmes au travail ont été votées. Au fil du temps, la femme, toujours en charge des tâches domestiques et de l'éducation des enfants pendant que leur mari travaillait à l'extérieur pour ramener de l'argent, a de plus en plus marqué sa place sur le marché du travail grâce notamment à la hausse de son niveau d'éducation et à la révolution industrielle. Un autre facteur important de la discrimination salariale vis-à-vis des femmes est *le choix d'orientation à l'école*. Des études sociologiques ont montré que les enseignants inculquent des aptitudes différentes chez les filles et les garçons. Concrètement, ils instaurent l'esprit de compétition auprès des garçons et le respect des règles auprès des filles ce qui entraîne que les filles éprouvent davantage de difficultés à choisir leur orientation. Or, le choix de l'orientation aura un impact sur la répartition des emplois entre les femmes et les hommes et, par conséquent, sur les inégalités salariales. Un autre facteur à prendre en compte est l'"*interaction entre la sphère domestique et la sphère productive*"³. Au cours de leur carrière, les femmes sont en effet amenées à faire des choix en termes d'interruptions de carrière, de régime de travail (travail à temps plein ou à temps partiel). Ces choix sont souvent liés à la présence d'enfants dans le ménage et ont un impact tant sur la rémunération des femmes que sur le comportement des embaucheurs. Bien que les femmes occupent une place croissante sur le marché du travail, la majorité d'entre elles restent les garantes des tâches ménagères et de l'éducation des enfants. Dans ce cadre, il est facile de comprendre la prédominance des femmes dans certains métiers tels qu'infirmières, enseignantes, techniciennes de surface, puéricultrices, aides à domicile.

¹ Le terme discrimination salariale doit être compris dans le sens: "*à travail égal, rémunération inégale*"

² DE CURRAIZE Y. et HUGOUNENG R. (2004), p. 197

³ DE CURRAIZE Y. et HUGOUNENG R. (2004), p. 193

Dans ce travail, nous tentons de comprendre: "***D'où viennent les inégalités de salaires entre hommes et femmes en Belgique?***". Bien que le principe de l'égalité de salaire entre les hommes et les femmes pour un même travail ait été inscrit en 1951 dans une Convention de l'Organisation Internationale du Travail, de nombreuses études montrent que ce principe n'est pas toujours respecté. Il nous paraissait donc important de comprendre d'où peuvent provenir ces inégalités salariales entre les deux sexes. De plus, ce sujet semble intéresser bon nombre de personnes, tels que les politiciens, l'Institut pour l'Egalité des Femmes et des Hommes, les chercheurs ou encore les particuliers.

Pour cela, nous disposons de plusieurs bases de données portant sur les années 2006 à 2010 reprenant 690 variables différentes. Celles-ci ont été mises à notre disposition par l'University of Amsterdam/Amsterdam Institute for Advanced Labour Studies dont les auteurs sont Tijdens, van Zijl, Hughie-Williams, van Klaveren, Steinmetz (2010). Nous nous sommes toutefois limités à l'année 2009 car il s'agissait de la base de données la plus récente et la plus complète du point de vue du nombre d'observations. Cette base de données a été établie à Amsterdam (Pays-Bas) en décembre 2010. Le modèle économétrique a été estimé pour ces données par la méthode des moindres carrés ordinaires (MCO).

Par ce modèle, nous avons tenté d'expliquer l'inégalité de salaire, la variable expliquée étant le salaire horaire, en fonction des variables explicatives qui sont le sexe, l'âge, le niveau d'éducation, l'expérience (le nombre d'années de services), le nombre d'enfants à charge, le type de contrat de travail (contrat à durée indéterminée ou pas), le statut d'emploi (employé, étudiant, retraité,...), le régime de travail (travail à temps plein ou à temps partiel), le statut marital, le type de secteur (public ou privé), le pourcentage de femmes en place dans l'entreprise, la position hiérarchique.

Le premier chapitre de ce travail présente une revue de la littérature sur le sujet et les résultats qui ont été obtenus.

Le deuxième chapitre montre notre démarche empirique. Nous expliquons notamment les caractéristiques des données, quelles sont les variables sélectionnées parmi les 690 variables de départ et quelle méthode a été utilisée.

Le troisième chapitre expose les résultats qui ont été obtenus, par une analyse descriptive et une analyse empirique.

Nous terminons par une conclusion générale qui reprend des éléments de réponse à la question de départ ainsi que les limites du modèle utilisé.

1. Revue de la littérature

A partir de 1957, Gary Becker, économiste américain, a dirigé plusieurs travaux sur la discrimination salariale. Selon lui, "*le salaire est fixé par les conditions de la concurrence et mesure la productivité qui est elle-même déterminée par l'ensemble des variables de capital humain des individus*"⁴ (préférences, aptitudes, parcours scolaire, expérience professionnelle, âge,...). "*S'il n'y a pas de discrimination, la différence salariale mesure la différence de productivité ou de caractéristiques individuelles liées à la production*"⁵. Dès lors, deux personnes ayant le même niveau de productivité devrait avoir la même rémunération. En d'autres termes, la différence salariale relèverait donc d'une différence de productivité due aux caractéristiques et choix des personnes. Toutefois, il faudrait être dans un monde où la discrimination n'existe pas pour pouvoir mesurer la part de la différence salariale due à la discrimination en tant que telle et celle due à aux caractéristiques et choix individuels.

Plusieurs études économétriques ont également été menées sur le sujet, que ce soit par Mincer et Polachek (1974) aux Etats-Unis, Thiry (1985), Glaude (1987), Sofer (1990), Lhéritier (1992), Bayet (1996) ou encore Meurs et Ponthieux (2000) en France⁶. Dans l'annexe 1, sont repris deux tableaux présentant ces études économétriques.

Ces études économétriques se sont basées sur des observations qui étaient à disposition à une période donnée, les données étaient dès lors exogènes. Elles ne tenaient pas compte des préférences, de la discrimination à l'embauche, ... qui sont difficilement quantifiables mais uniquement de la discrimination salariale. La discrimination salariale signifiant une différence de salaire entre des personnes ayant la même fonction ou les mêmes caractéristiques. L'estimation de celle-ci se basait sur la décomposition d'Oaxaca-Blinder de 1973 dont l'explication détaillée est reprise au point 2.2.2. Selon cette méthode, l'écart salariale proviendrait soit de la discrimination, soit de caractéristiques personnelles différentes constatées tels que l'expérience professionnelle, l'âge, le niveau d'éducation, ... La méthode d'Oaxaca-Blinder avait cependant des limites tels que le biais de sélection dans l'échantillon observé dû à un moindre pourcentage de participation pour les femmes. Ponthieux et Meurs (2000) ont tenté de remédier à ce problème de sélection. Une autre limite de la méthode était que la mesure de l'estimation des équations de salaires était uniquement basée sur des variables personnelles tels que l'âge, l'expérience professionnelle, le niveau d'éducation, ... mais il s'est avéré que tenir compte de ces variables seulement étaient insuffisant pour expliquer la formation des rémunérations. En effet, un facteur à prendre en compte serait l'hétérogénéité des entreprises car les rémunérations diffèrent d'une entreprise à l'autre, d'un secteur d'activité à l'autre. La prise en compte de l'hétérogénéité des entreprises de même que le taux de féminisation dans l'entreprise ou le secteur diminueraient la mesure de la discrimination salariale.

La décomposition d'Oaxaca-Blinder a aussi été développée par Brown, Moon et Zoloth (1980). Ils y ont intégré la probabilité plus faible pour les femmes que pour les hommes d'atteindre certaines catégories d'emploi ou niveaux hiérarchiques. Cette méthode a toutefois été peu utilisée faute de données.

⁴ DE CURRAIZE Y. et HUGOUNENG R. (2004), p. 199

⁵ DE CURRAIZE Y. et HUGOUNENG R. (2004), p. 199

⁶ DE CURRAIZE Y. et HUGOUNENG R. (2004), pp. 206-209

Les graphiques repris ci-dessous ont pour objectif de montrer la part de l'écart salarial inexplicable par le modèle étudié et qui est imputable à la discrimination salariale et ce selon l'ethnie d'une part et en fonction du statut marital d'autre part. Les données qui nous ont permis d'élaborer ces graphiques sont issues des tableaux repris dans l'annexe 1.

Graphique 1

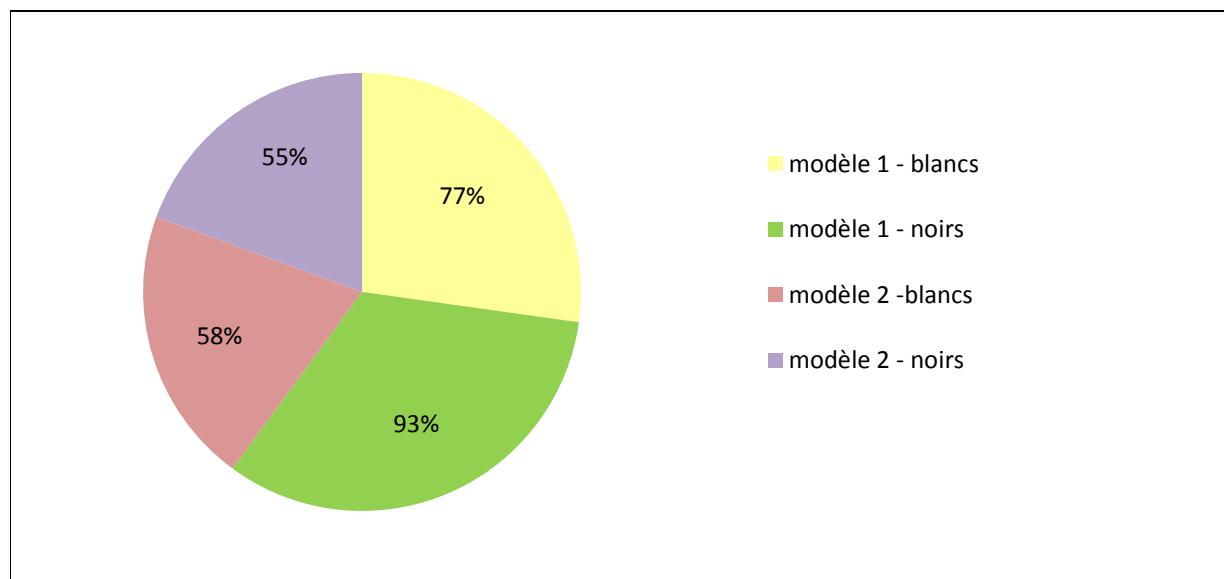
Dans le cas du **modèle 1** de l'étude menée par Oaxaca (1973), les variables explicatives suivantes: expérience, éducation, état de santé, régime de travail, migration, statut marital, taille de la ville et région sont prises en compte dans l'équation de salaire. Le champ d'application portait sur les personnes de couleur blanche et celles de couleur noire.

Le graphique montre que pour les personnes de couleur blanche, 77% du différentiel de salaire inexplicable par le modèle peut être attribué à la discrimination salariale. Pour les personnes de couleur noire, ce pourcentage est plus élevé, soit 93%.

Dans le cas du **modèle 2**, le type d'occupation et le secteur d'activité ajoutés aux paramètres du modèle 1 sont pris en considération dans l'équation de salaire. Ce modèle ne tient pas compte de la discrimination à l'embauche.

Le graphique montre dans ce cas que, pour les personnes de couleur blanche, 58% du différentiel de salaire inexplicable par le modèle peut être imputable à la discrimination salariale contre 55% pour les personnes de couleur noire.

Graphique 1: Part de l'écart salarial inexplicable par le modèle et imputable à la discrimination salariale selon l'ethnie (en %)



Source: auteurs sur base de l'annexe 1

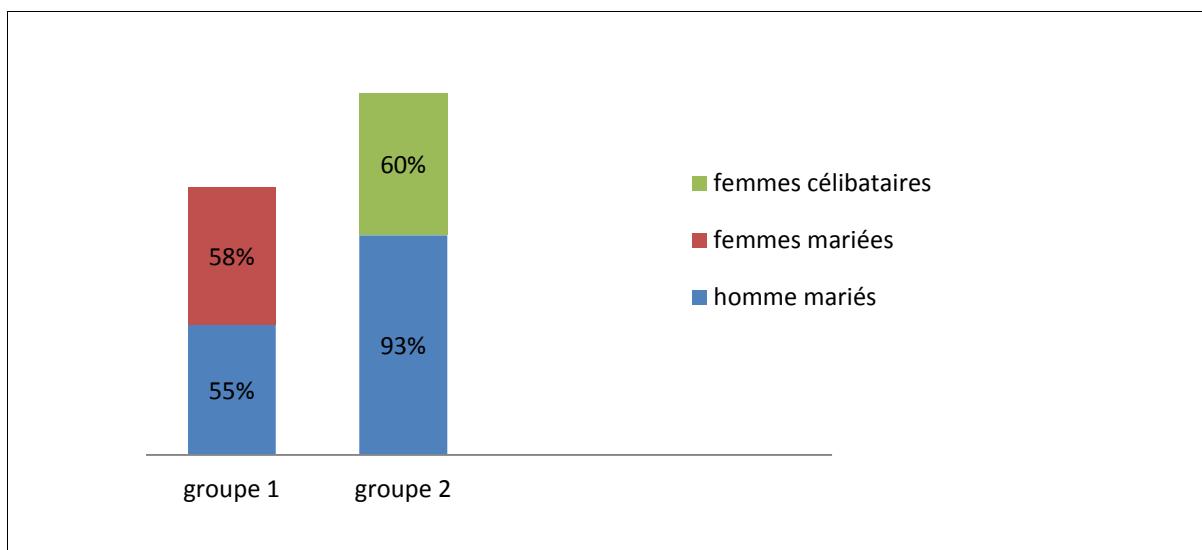
Dans le modèle 1, on constate qu'une importante part de l'écart salarial inexplicable par le modèle peut être imputable à la discrimination salariale tandis que dans le cas du modèle 2, cette part de l'écart salarial inexplicable est moindre. Dès lors, tenir compte du type d'occupation et du secteur d'activité ainsi que l'absence de discrimination à l'embauche diminue de façon assez considérable la part de l'écart salarial inexplicable par le modèle et imputable à la discrimination. Cette diminution est valable tant pour les personnes blanches de peau que pour les personnes noires de peau.

Graphique 2

Dans l'étude menée par Mincer et Polacheck (1974), les variables explicatives suivantes: expérience et niveau d'éducation sont prises en compte dans l'équation de salaire; le champ d'application étant le statut marital. Le **groupe 1** comporte les hommes mariés et les femmes mariées; le **groupe 2** est composé des hommes mariés et des femmes célibataires.

Le graphique montre que la part de l'écart de salaire inexplicable par le modèle et attribuable à la discrimination salariale est comprise entre 55% et 58% pour la comparaison entre hommes mariés et femmes mariées et entre 60% et 93% pour la comparaison entre hommes mariés et femmes célibataires.

Graphique 2: Part de l'écart salarial inexplicable par le modèle et imputable à la discrimination salariale selon le statut marital (en %)



Source: auteurs sur base de l'annexe 1

Il est dès lors évident que, dans le cas du groupe 2, la part de l'écart salarial inexplicable par le modèle et imputable à la discrimination est fortement plus élevée que dans le groupe 1. Le statut marital a donc bel et bien un impact.

Comme le montrent les graphiques, la différence de salaire peut être expliquée par une part significative de discrimination salariale. Toutefois, il est difficile de mesurer son ampleur car les préférences, par exemple, difficiles à mesurer, pourraient également expliquer la différence de rémunération. Par ailleurs, la prise en compte de variables d'emploi (taux de syndicalisation,...) en plus des variables individuelles (âge, niveau d'éducation,...) dans l'équation de salaire réduirait la part de la discrimination mesurée. Néanmoins, le manque d'information sur les variables d'emploi et la priorité mise sur la théorie du capital humain entraînent que ce type de variables ne sont généralement pas prises en compte dans l'équation de salaire. En conclusion, même si la part de l'écart de salaire inexplicable par un modèle peut être imputée à la discrimination salariale, il faut néanmoins garder en mémoire que la discrimination salariale n'expliquerait qu'une partie de ce différentiel de salaire inexplicable.

2. Démarche Empirique

2.1. Base de données

2.1.1. La base de données

Une explication quant à la source des données est reprise en annexe 2.

Dans le cadre de notre étude, nous avons reçu quatre bases de données reprenant les années 2006, 2007, 2008 et 2009. Les données avaient été générées par SSPH18, que nous avons alors transformé par le programme « Stat transfert » afin de pouvoir les lire et les utiliser avec « XLStat ».

Nous avons conservé l'année 2009, celle qui était la plus récente et la mieux pourvue en caractéristiques ainsi qu'en échantillons.

Au départ, nous disposions d'une base de données brute comprenant 691 colonnes et 11.468 enregistrements.

Nous avons sélectionné les variables qui nous semblaient les plus pertinentes pour l'étude de notre modèle et, par conséquent, n'avons conservé que 13 colonnes dont le détail des variables est repris ci-après.

Etant donné qu'il nous semblait inopportun de travailler sur des échantillons dont nous ne disposions pas de tous les renseignements, nous avons, dans un premier temps, filtré et nettoyé la base de données comme suit:

Tableau 1: Nettoyage de la base de données

FILTRE	VARIABLE	NOMBRE TOTAL D'ENREGISTREMENTS	CELLULES VIDES	NOMBRE RESTANT D'ENREGISTREMENTS APRES SUPPRESSION DES CELLULES VIDES
1	WAGEHRPP	11.468	8.190	3.278
2	GENDER	3.278	680	2.598
3	TENUEXPE	2.598	30	2.598
4	EDUCATYR	2.568	215	2.353
5	CONTR7	2.353	150	2.203
6	chld	2.203	10	2.193
7	FIRMPRI	2.203	46	2.147

Note 1: Ces données sont issues de l'étude de Tijdens, K.G. et Osse P., "WageIndicator continuous web-survey on work and wages." Amsterdam: University of Amsterdam/AIAS and StichtingLoonwijzer. (WAGEHRPP) représente le salaire horaire brut, (GENDER) le sexe de l'individu, (TENUEXPE) le nombre d'année de service, (EDUCATYR) le degré de scolarité de l'individu, (CONTR7) le type de durée de contrat, (chld) si l'individu a des enfants ou non, (FIRMPRI) si l'individu est employé dans une firme privée ou publique.

Il nous restait alors pour notre étude des données en coupe complètes concernant 2.147 hommes et femmes belges pour l'année 2009.

2.1.2. Les caractéristiques des données

Les caractéristiques intrinsèques de l'individu sont reprises dans cette étude avec la variable "dummy" qui représente le **sexe de l'individu** (Gender) et qui prend la valeur 1 si la personne interrogée est une femme et 0 s'il s'agit d'un homme, la variable (AGE) représentant **son âge au moment de l'étude**, le **nombre d'années d'étude** poursuivie après les études primaires (EDUCATYR). Le fait que l'individu ait ou non des **enfants** (chld) est représenté par une variable "dummy" qui prend la valeur 1 dans l'affirmative et 0 dans la négative. L'individu ayant un **contrat à durée indéterminée** (CONTR7) est également caractérisé par une variable binaire valant 1 si oui, sinon 0.

D'autres variables binaires ont été construites pour permettre l'analyse en tenant compte d'un maximum de facteurs intervenant dans la formation des salaires. La variable (CONSTRO) représentant le **statut courant de l'emploi** a été fractionnée en quatre sous-groupes chacun traduit par une variable "dummy" suivant la branche de l'emploi de l'individu.

Il en va de même pour la variable (firmfema), **pourcentage de femmes dans l'entreprise**. Dans ce cas-ci, nous avons dénombré 5 fourchettes de pourcentage, les variables binaires associées sont générées également pour créer 5 variables indépendantes.

Le **secteur d'activité étudié** (FIRMPRI) a lui aussi été exprimé en 4 types de secteurs relevés dans l'étude par des variables "dummy" indépendantes.

La variable relative au **statut marital** (HHSTA) est décomposée en 4 statuts distincts reprenant les diverses possibilités maritales par des variables "dummy" propres à chaque type.

La **position occupée dans la hiérarchie de l'entreprise** (occtype) a été agrégée suivant le rang en 4 catégories qui sont chacune également révélées par des variables "dummy".

Enfin, [Ln(WAGEHRPP)] représente le **logarithme népérien du salaire horaire brut des individus en 2009**. Il s'agit d'une variable construite à partir du **salaire horaire brut** (WAGEHRPP) permettant d'étudier les variations du salaire en pourcentage.

Toutes ces variables sont reprises en détail dans le tableau ci-après.

Tableau 2: Liste des variables et significations

VARIABLE NAME	VARIABLE LABEL
AGE	Age
chld	Has children
CONTR7	Has permanent employment contract
CONTSTRO	Current employment status - used for routing
CONTSTRO1	Employee
CONTSTRO1	Civil servant Employee
CONTSTRO1	Paid family worker/working for family business
CONTSTRO1	In a job creation scheme
CONTSTRO1	In military or comparable services
CONTSTRO7	Student in full-time education Student
CONTSTRO8	Apprentice
CONTSTRO9	Retired, previously no job Other
EDUCATYR	Length of education in years (excluding pre-school)
firmfema	Percentage women in workplace
firmfema0-20%	Percentage women in workplace 0 to 20%
firmfema20-40%	Percentage women in workplace 20 to 40%
firmfema40-60%	Percentage women in workplace 40 to 60%
firmfema60-80%	Percentage women in workplace 60 to 80%
firmfema80-100%	Percentage women in workplace 80 to 100%
FIRMPRI	Public or private sector
FIRMPRI1	Private sector
FIRMPRI2	Public sector
FIRMPRI3	Non-profit sector
FIRMPRI4	No specification for sector
GENDER	Gender
HHSTAT	Current marital status
HHSTAT1	never married
HHSTAT2	married
HHSTAT3	widowed
HHSTAT4	divorced
HRSHISTO	Has full-time working hours (self-defined)
occtype	Position in occupational hierarchy
occtype1	Hierarchy occupational pool category 1 1 'No, the job title is alright'
occtype2	Hierarchy occupational pool category 2 to 6 2 'Assistant' 3 'Junior' 4 'Apprentice' 5 'Trainee' 6 'Trainer, training responsibilities'
occtype3	Hierarchy occupational pool category 7 to 8 7 'Foreman, forewoman' 8 'Senior'
occtype4	Hierarchy occupational pool category 9 to 12 9 'Team leader' 10 'Supervisor' 11 'Head' 12 'Master'
TENUEXPE	Years of service
Ln(WAGEHRPP)	Log népérien Wage per hour (gross) controlled for PPP

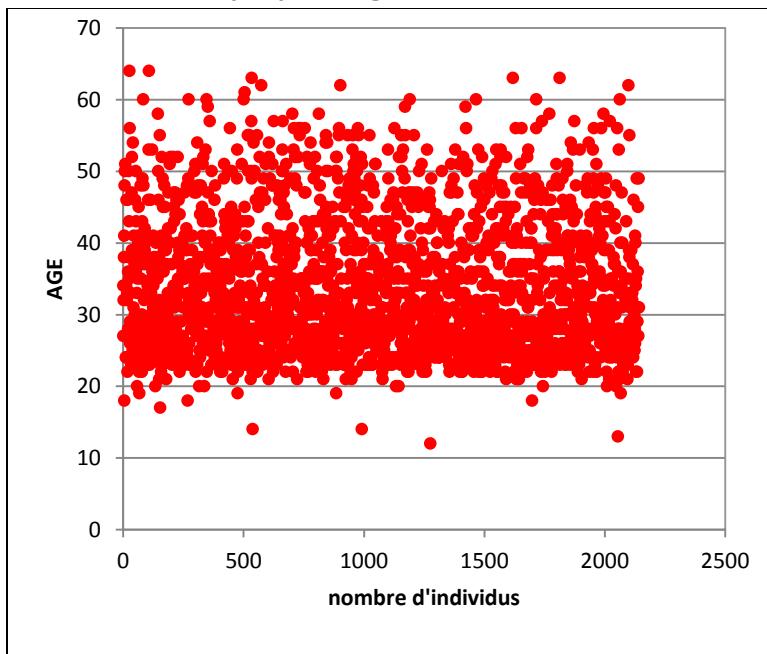
Nous avons entrepris l'analyse des données en procédant en premier par une étude statistique et graphique afin de préparer au mieux l'analyse économétrique.

Cette étape a permis, d'une part, de vérifier que la base de données ne comportait pas de valeurs manquantes ou aberrantes et, d'autre part, de refléter l'existence d'un lien entre le logarithme du salaire et ses déterminants.

Les différents graphiques en annexe nous ont permis de pouvoir éliminer certaines aberrations dans la base de données.

Le Graphique 3 relatif à l'âge des individus nous indique en l'occurrence que 5 valeurs se trouvent en dessous de l'âge minimal pour prétendre à pouvoir travailler en Belgique. Ces valeurs seront supprimées de notre étude.

Graphique 3: Age des Individus



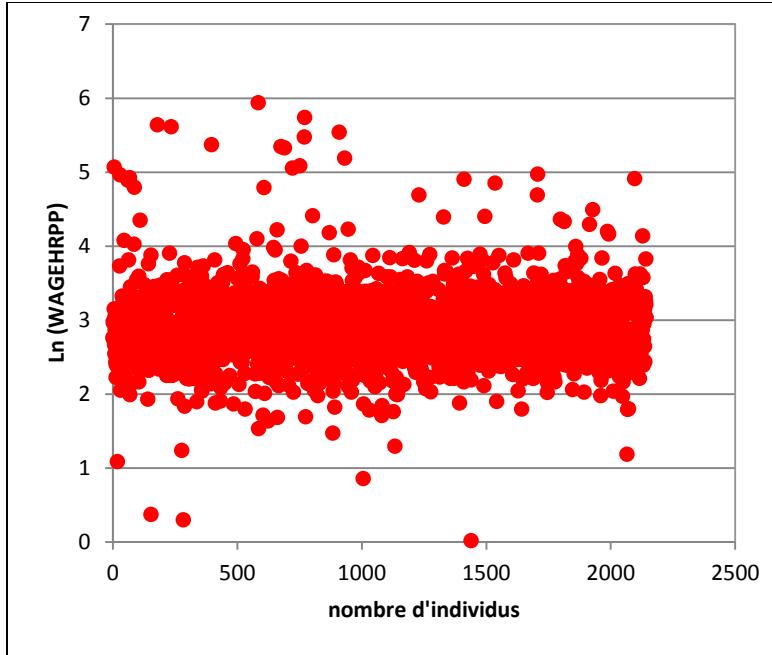
Source: *Tijdens et Osse (2010)*

Tableau 3: Données supprimées

	DATA SUPPRIMÉE	AGE CORRESPONDANT
1	154	17 ans
2	538	14 ans
3	992	14 ans
4	1277	12 ans
5	2056	13 ans

Le Graphique 6 correspondant aux logarithmes népériens des salaires horaires bruts des individus nous a permis plus facilement de visualiser la data 1440. Cette data nous indiquait que le salaire de l'individu était de 1,015 euro brut de l'heure, valeur qui est supprimée également de la base de données.

Graphique 6: Logarithme népérien du salaire horaire brut des individus



Source: Tijdens et Osse (2010)

Tableau 4: Statistiques descriptives

	MOYENNE	ECART-TYPE D'ECHANTILLON	MINIMUM	MAXIMUM
AGE	33,985	9,307	18,000	64,000
EDUCATYR	13,826	2,148	5,000	16,000
TENUEXPE	11,756	9,738	0,000	46,000
Ln (WAGEHRPP)	2,889	0,485	0,296	5,933

Note 2: Statistiques descriptives des séries de logarithme du salaire, de l'âge de l'individu, du nombre d'années d'étude, de l'ancienneté dans l'emploi actuel pour 2.137 individus belges en 2009.

Notre base de données nettoyée concernant les déterminants du salaire des individus belges en 2009 comporte « in fine » 2.137 observations pour 2.147 au départ de l'analyse. En 2009, le logarithme du salaire était compris entre 0,296 et 5,933 pour une moyenne de 2,889, ce qui correspond en passant par son exponentielle à un salaire horaire brut de 17,98 euros. L'âge moyen des individus repris dans cette étude est de 34 ans, ce qui représente un échantillon de population plutôt jeune. Le nombre d'années d'étude moyen est de 14 ans, tandis que l'expérience moyenne est de 11,76 ans.

2.1.3. Lien entre la variable que l'on cherche à expliquer et les variables explicatives

Les liens entre les variables ont des coefficients de corrélation linéaires de Bravais-Pearson compris entre 0,02 et 0,326 pour le log du salaire et ses déterminants potentiels. Les corrélations sont positives pour l'âge (0,326), l'expérience dans le milieu du travail (0,285), le fait d'avoir des enfants (0,215), la durée de la scolarité (0,216), le statut marital de l'individu (0,211), également s'il a un contrat à durée déterminée (0,143), l'occupation hiérarchique plus élevée de l'individu dans la société (0,141 occtype3 et 0,247 occtype4).

2.1.4. Lien entre les variables explicatives

En ce qui concerne la corrélation des déterminants entre eux, les chiffres oscillent entre -0,07 et 0,326. Ceci exclut quasiment des cas de multicolinéarité parfaite entre variables. La matrice de corrélation est reprise à l'annexe 7.

2.2. Méthode utilisée

2.2.1. La méthode des Moindres Carrés Ordinaires

Il s'agit de la méthode d'estimation que nous avons utilisée dans le cadre de ce travail car, comme explicité ci-après, elle nous permet d'analyser les déterminants du logarithme du salaire. Le but de cette méthode est de trouver les paramètres qui minimisent la somme des carrés des erreurs/résidus.

Toutefois, la qualité des estimateurs dépend des hypothèses sur lesquelles reposent la technique et la qualité des données. Pour que les estimateurs soient de bonne qualité, ils doivent vérifier certaines hypothèses:

- H1: espérance nulle du terme d'erreur
- H2: indépendance entre les variables explicatives et le terme d'erreur
- H3: variation des variables dans le temps
- H4: absence d'autocorrélation entre les termes d'erreur et homoscédasticité des erreurs, c'est-à-dire variance constante sur l'échantillon.
- H5: stationnarité⁷
- H6: terme d'erreur suit une loi normale d'espérance nulle et de variance σ^2

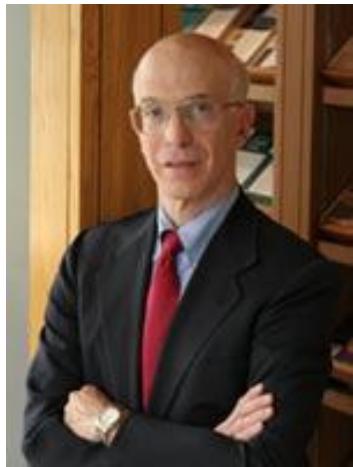
C'est donc sur base de ces hypothèses que l'on pourra déterminer la qualité des estimateurs. Nous y reviendrons notamment en page 17 et 18.

⁷ L'hypothèse 5 n'a pas été vue au cours

2.2.2. La méthode d'Oaxaca-Blinder (1973)



Ronald L. OAXACA



Alan S. BLINDER

Certains économistes se sont lancés dans la recherche de nouveaux outils économétriques qui permettraient de mesurer la portion expliquée de la différence des salaires, c'est-à-dire attribuable à des dotations différentes de capital humain entre hommes et femmes et celle qui reste inexpliquée.

La technique de décomposition d'Oaxaca-Blinder est le fruit du travail de deux économistes, Alan Blinder et Ronald Oaxaca, qui l'introduisirent en 1973 dans la littérature économique.

Une version similaire de cette technique avait déjà été utilisée en sociologie vers la fin des années 1960 et début des années 1970 afin d'étudier l'origine des écarts de salaires raciaux (Duncan 1969, Althauser et Wigler 1972).

Cette méthode est devenue l'outil de base pour étudier les différences salariales entre les sexes ainsi que les discriminations raciales. De plus, elle a été admise auprès des tribunaux aux Etats-Unis dans les procédures judiciaires relatives aux discriminations (Ashenfelter et Ronald, 1987).

Elle permet de mesurer la portion expliquée de la différence des salaires, c'est-à-dire attribuable à des dotations différentes de capital humain entre homme et femme et celle qui reste inexpliquée, donc d'estimer la discrimination en décomposant la différence totale de salaire entre les femmes et les hommes.

Si le mode de formation des salaires des hommes et des femmes est différent, il faut estimer les équations de ces deux groupes séparément.

On peut ainsi écrire les deux régressions comme suit :

$$- \text{Pour les hommes : } \ln W_i^m = \alpha_m + \beta_m X_i + \varepsilon_i^m \quad (1)$$

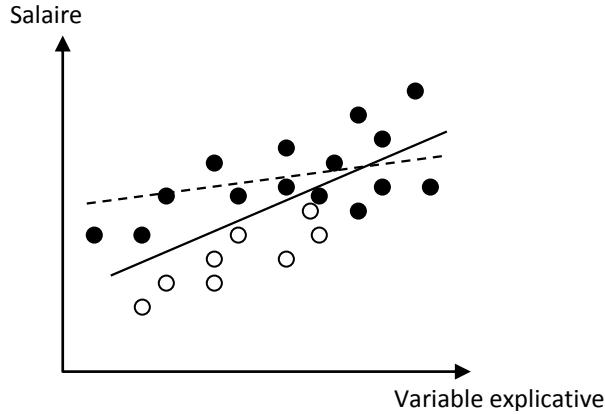
$$- \text{Pour les femmes : } \ln W_i^f = \alpha_f + \beta_f X_i + \varepsilon_i^f \quad (2)$$

α et β étant des paramètres pouvant être estimé par les Moindres Carrés Ordinaires (MCO).

Etant donné que la droite de régression des MCO passe par le point moyen:

$$\overline{\ln W_m} = \hat{\alpha}_m + \hat{\beta}_m \cdot \bar{X}_m$$

$$\overline{\ln W_f} = \hat{\alpha}_f + \hat{\beta}_f \cdot \bar{X}_f$$



la différence entre le salaire moyen des hommes et des femmes s'écrira:

$$\overline{\ln W_m} - \overline{\ln W_f} = \hat{\alpha}_m - \hat{\alpha}_f + \hat{\beta}_m \cdot \bar{X}_m - \hat{\beta}_f \cdot \bar{X}_f$$

En ajoutant dans un premier temps: $\hat{\beta}_f \cdot \bar{X}_m - \hat{\beta}_f \cdot \bar{X}_m$

On obtient l'équation suivante qui mesure la *discrimination des hommes*:

$$\overline{\ln W_m} - \overline{\ln W_f} = \hat{\alpha}_m - \hat{\alpha}_f + (\hat{\beta}_m - \hat{\beta}_f) \cdot \bar{X}_m + \hat{\beta}_f (\bar{X}_m - \bar{X}_f)$$

En ajoutant dans un deuxième temps: $\hat{\beta}_m \cdot \bar{X}_f - \hat{\beta}_m \cdot \bar{X}_f$

On obtient l'équation suivante qui mesure la *discrimination des femmes*:

$$\overline{\ln W_m} - \overline{\ln W_f} = \hat{\alpha}_m - \hat{\alpha}_f + (\hat{\beta}_m - \hat{\beta}_f) \cdot \bar{X}_f + \hat{\beta}_m (\bar{X}_m - \bar{X}_f)$$

On suppose implicitement que l'équation salariale des hommes constitue la valorisation normale que tout le marché du travail devrait accorder à toutes personnes possédant ces qualités.

3. Compte rendu de l'analyse

3.1. Analyse descriptive

Dans cette partie, nous avons procédé à l'analyse économétrique des déterminants du salaire. Le modèle économétrique final que l'on souhaite estimer est le modèle linéaire multiple suivant :

Modèle (1)

$$\begin{aligned} \ln W_i = & \alpha + \beta_1 \text{AGE} + \beta_2 \text{chld} + \beta_3 \text{CONTR7} + \beta_4 \text{CONTSTRO} + \beta_5 \text{EDUCATYR} \\ & + \beta_6 \text{firmfema} + \beta_7 \text{FIRMPRI} + \beta_8 \text{GENDER} + \beta_9 \text{HHSTAT} \\ & + \beta_{10} \text{HRSHISTO} + \beta_{11} \text{occtype} + \beta_{12} \text{TENUEXPE} + \varepsilon_i \end{aligned} \quad (i = 1, \dots, 2143)$$

Avec comme variable endogène **ln W_i** , le logarithme du salaire de l'individu i , et comme variables exogènes, **chld** le fait d'avoir des enfants, **CONTR7** s'il a un contrat permanent, **CONTSTRO** pour le type d'emploi, **EDUCATYR** le nombre d'années d'étude complétées en 2009, **firmfema** étant le pourcentage de femmes dans l'entreprise, **FIRMPRI** indiquant le secteur d'activité, **GENDER** étant le sexe de l'individu (0 pour les hommes et 1 pour les femmes), **HHSTAT** pour le statut marital de l'individu, **HRSHISTO** s'il est à temps plein ou non, **occtype** représente la position hiérarchique de l'individu dans l'entreprise, et pour finir, **TENUEXPE**, l'ancienneté dans le poste actuel. Tous ces paramètres sont les paramètres inconnus que l'on cherche à estimer.

Enfin, ε_i représente le terme d'erreur et i le nombre d'observations. Nous avons énuméré les différentes variables binaires construites liées à la création de catégories plus affinées, elles sont reprises dans le tableau 2 (page 11) reprenant toutes les variables utilisées dans cette étude.

Cette spécification correspond à un **modèle linéaire multiple** qui nous permet de limiter le biais d'omission de variables. Nous avons utilisé la **méthode d'estimation des moindres carrés ordinaires**.

Cette méthode nous permet d'analyser les déterminants du logarithme du salaire car elle fournit sous les hypothèses de Gauss-Markov discutées ci-dessous une estimation non biaisée des paramètres inconnus avec une variance minimale, i.e. une bonne approximation des paramètres inconnus. Dans le paragraphe qui suit, nous allons discuter des hypothèses du modèle. Certaines seront assumées tandis que d'autres seront explicitement testées.

On supposera ici certaines hypothèses du modèle linéaire classique vérifiées (hypothèses de Gauss-Markov, plus hypothèse de normalité du terme d'erreur), d'autres seront explicitement testées. On suppose notamment que les perturbations ont une espérance nulle⁸ (H1) du fait de l'introduction de la constante dans notre modèle et que les régresseurs sont strictement exogènes (H2).

⁸ Nous pouvons vérifier cette hypothèse à postériori avec le logiciel XLStat afin de vérifier si les résidus ont bien une espérance nulle.

En d'autres termes, nous supposons que les régresseurs utilisés ici qui représentent les caractéristiques des individus sont indépendants du terme d'erreur. L'hypothèse H3 est vérifiée dans notre échantillon car la variance empirique de nos régresseurs est différente de zéro.

L'hypothèse H4 d'absence de colinéarité parfaite semble également vérifiée suite à l'analyse des coefficients de corrélation linéaire. Il reste deux hypothèses dans le modèle linéaire classique :

1. l'homoscédasticité
2. la normalité des termes d'erreur

Nous proposons ci-dessous de les tester formellement.

A l'annexe 4, vous pourrez trouver également un exemple de l'application du test de la normalité des résidus de Jarque-Bera pour le modèle (1).

Ce test basé sur les moments d'ordre 3 et 4 de la loi normale (*skewness et kurtosis*) qui valent respectivement 0 et 3 pour la loi normale.

Ici, le test rejette l'hypothèse H0 de normalité des résidus. En général, les tests de normalité des résidus rejettent l'hypothèse nulle du fait de la présence de valeurs extrêmes qu'il convient de traiter avec des variables « dummies ».

Il est délicat d'interpréter économiquement les valeurs extrêmes et nous utilisons les résultats du théorème centrale-limite pour poursuivre notre étude, à savoir que le fait d'avoir un nombre élevé d'observations (2137) permet de supposer que le terme d'erreur est normalement distribué.

Nous effectuons l'estimation avec le logiciel XLStat pour le premier modèle en ne conservant que les coefficients significativement différent de zéro au seuil de 5% dont les résultats sont présentés à l'annexe 5.

3.2. Analyse empirique

3.2.1. Régression directe sur les deux groupes hommes-femmes

Tableau 5: Résultats du modèle linéaire multiple

Source	Valeur	Ecart-type	t	Pr > t
Constante	1,692	0,156	10,861	< 0,0001
AGE	0,012	0,003	3,877	0,000
chld	0,033	0,025	1,345	0,179
CONTR7	0,044	0,035	1,238	0,216
CONTSTRO1	-0,110	0,122	-0,896	0,370
CONTSTRO7	0,136	0,178	0,763	0,446
CONTSTRO8	0,039	0,272	0,142	0,887
CONTSTRO9	0,000	0,000		
EDUCATYR	0,061	0,005	13,042	< 0,0001
firmfema0-20	0,038	0,038	1,001	0,317
firmfema20-40	0,100	0,038	2,597	0,009
firmfema40-60	0,110	0,037	2,984	0,003
firmfema60-80	0,061	0,041	1,498	0,134
firmfema80-100	0,000	0,000		
FIRMPRI1	0,025	0,019	1,283	0,200
FIRMPRI2	-0,059	0,040	-1,499	0,134
FIRMPRI3	-0,012	0,043	-0,285	0,776
FIRMPRIvide	0,000	0,000		
GENDER	-0,096	0,021	-4,638	< 0,0001
HHSTAT1	0,073	0,023	3,120	0,002
HHSTAT2	0,258	0,191	1,350	0,177
HHSTAT3	-0,058	0,044	-1,308	0,191
HHSTAT4	0,000	0,000		
HRSHISTO	0,075	0,032	2,318	0,021
Occtype1	-0,203	0,025	-8,132	< 0,0001
Occtype2	-0,272	0,031	-8,836	< 0,0001
Occtype3	-0,095	0,033	-2,861	0,004
Occtype4	0,000	0,000		
TENUEXPE	0,004	0,003	1,236	0,217
R ²	0,270	-	-	-
Observations	2137	-	F	Pr > F
Fisher	-	-	33,979	< 0,0001

Note 3: Les résultats proviennent de l'estimation d'un modèle M.C.O. Le logiciel utilisé est XLStat. La variable expliquée est le logarithme du salaire horaire brut des individus, les variables explicatives sont pour chaque individus : l'âge, le fait qu'il ait ou non des enfants, un contrat à durée déterminée ou non, la position de son emploi dans la hiérarchie de l'entreprise, la durée de ses études, le pourcentage de femmes dans l'entreprise, si l'entreprise relève du secteur privé ou public, le sexe de l'individu, son statut marital, son expérience dans le milieu du travail.

Qualité du modèle:

Les résultats du premier modèle permettent de déterminer les variables qui ont un impact significatif sur les salaires horaires bruts. La question de départ vise à déterminer si le sexe des individus est prépondérant dans les écarts de salaires qui pourraient être décelés, et de ce fait, mettre en évidence une discrimination entre le salaire des hommes et celui des femmes.

Le test de Fisher, repris à l'annexe 6, permet de réaliser un test joint sur tous les coefficients du modèle. Pour le modèle (1), la valeur de la F-stat est de 40 et sa P-value est de 0,0001% ce qui nous permet de vérifier qu'il existe au moins un coefficient significativement différent de zéro au sein du modèle (1).

Tests individuels:

Afin de pouvoir détecter dans un premier temps quelles variables pouvaient avoir une action significative sur les salaires horaires bruts, nous avons procédé à des tests de Student bilatéraux (i.e. on ne fait pas d'hypothèse sur le signe du coefficient).

Ce test permet de vérifier la significativité des coefficients du modèle (1) et de ne retenir que les coefficients significativement différents de zéro (voir tableau 5 supra).

Nous effectuons ce test malgré l'absence de normalité du terme d'erreur en supposant que les résultats du théorème central limite s'appliquent bien (i.e. convergence asymptotique des distributions) de la façon suivante :

On teste :
$$\begin{cases} H_0: \beta_j = 0 \\ \text{Contre } H_a: \beta_j \neq 0 \end{cases}$$

Pour cela, on calcule la t-stat :

$$t - \text{stat} \equiv \frac{\hat{b}_j - b_{H_0}}{se(\hat{b}_j)}$$

Sous l'hypothèse nulle (i.e. si l'hypothèse nulle est vraie) cette statistique suit une distribution connue :

$$t - \text{stat} \sim t(n - k - 1)$$

où $t(n - k - 1)$ est une distribution de Student avec $(n - k - 1)$ degrés de liberté, n étant le nombre d'observations dans l'échantillon et k le nombre de régresseurs.

Les valeurs de t-stat sont présentées dans le tableau 5 ci-dessus. Le logiciel XLSTAT présente également les p-values correspondantes. Les p-value fournissent les probabilités de rejeter l'hypothèse nulle à tort. Pour procéder au test, il nous faut définir une règle de décision.

Si la p-valeur est inférieur à 5%, on décide de **rejeter l'hypothèse nulle**. A l'inverse, **on ne pourra rejeter l'hypothèse nulle** si la p-valeur est supérieure à 5%.

Dans le tableau 5 présenté en page 19, on constate que certains coefficients sont statistiquement significatifs au seuil de 5%.

Interprétation des résultats:

La régression générale sur l'ensemble des hommes et des femmes avec les 24 variables reprises dans le tableau ci-dessus permet d'expliquer 27% de la variation du logarithme des salaires horaires bruts (R^2 ajusté = 0,27), ce qui est faible pour une coupe transversale. L'éducation ($t = 13,042$) est l'attribut le plus significatif ce qui est logique, suivi de l'âge ($t = 3,877$) ce qui ne surprend pas puisque l'éducation est un corolaire de l'âge.

Le taux d'occupation de femmes dans l'entreprise, lorsque celui-ci est de 40 à 60% influence également sur le salaire ($t = 2,984$), de 20 à 40 % d'occupation ($t = 2,597$), le statut marital également ($t = 3,120$), le fait que l'entreprise soit du secteur privé ($t = 1,283$) et enfin, le fait d'avoir des enfants ($t = 1,345$), avoir un contrat à durée indéterminée ($t = 0,763$) et pour terminer l'expérience dans le travail ($t = 1,236$).

Le coefficient de la variable « GENDER » (0 pour les hommes et 1 pour les femmes) est négatif et significatif ($t = -4,638$) ce qui signifie que selon cette régression générale et directe, il y aurait une discrimination salariale nettement marquée envers les femmes « *ceteris paribus* », **les femmes auraient un salaire inférieur d'environ 9,6% par rapport à celui des hommes.**

Tableau 6: Résultats de la régression sur les paramètres significatifs

Paramètre	Valeur	Ecart-type	t de Student	Pr > t
Constante	1,638	0,085	19,206	< 0,0001
AGE	0,015	0,001	13,915	< 0,0001
EDUCATYR	0,059	0,004	13,445	< 0,0001
firmfema20-40	0,062	0,023	2,666	0,008
firmfema40-60	0,072	0,021	3,341	0,001
GENDER	-0,101	0,019	-5,193	< 0,0001
HHSTAT1	0,093	0,019	4,783	< 0,0001
HRSHISTO	0,060	0,032	1,908	0,057
Occtype1	-0,210	0,025	-8,457	< 0,0001
Occtype2	-0,277	0,030	-9,119	< 0,0001
Occtype3	-0,092	0,033	-2,808	0,005

Une interprétation plus ciblée sur les coefficients de la seconde régression, nous indique qu'une année d'âge supplémentaire, augmente le salaire de 1.5 %, une année d'étude supplémentaire de 5.9%, le fait de travailler dans une entreprise où les femmes sont représentatives pour 20-40 et 40-60 % entraîne une différence de salaire positive de 6.2 à 7.2 %.

Le statut marital, si la personne est mariée, le salaire est augmenté de 9.3 % par rapport à un célibataire ou autre. Le fait de travailler à temps plein, la différence est positive par rapport à un part-time, de l'ordre de 6%.

Pour les incidences négatives, **le sexe comme déjà constaté, entraîne une différence sur le salaire de 10.1% des femmes par rapport à celui des hommes**, la hiérarchie du travailleur dans l'entreprise ; si le travailleur n'a pas de qualification, la différence est de 21%, de 27.7% si le travailleur est occupé dans le pool des catégories de 2 à 6, et enfin de 9.2% si sa catégorie d'emploi est de 7-8 (voir tableau 2 en page 11).

3.2.2. Régression sur le groupe hommes

Tableau 7: Régression sur les hommes

Source	Valeur	Ecart-type	t	Pr > t
EDUCATYR	0,06	0,007	8,956	< 0,0001
Constante	1,675	0,291	5,764	< 0,0001
AGE	0,022	0,005	4,544	< 0,0001
HHSTAT1	0,083	0,036	2,278	0,023
chld	0,052	0,037	1,381	0,168
firmfema20-40	0,097	0,092	1,051	0,293
CONTSTRO8	0,383	0,389	0,983	0,326
firmfema40-60	0,085	0,092	0,918	0,359
HRSHISTO	0,063	0,091	0,687	0,492
FIRMPRI1	0,008	0,028	0,278	0,781
firmfema0-20	0,025	0,091	0,277	0,782
firmfema60-80	0,028	0,101	0,275	0,783
CONTR7	0,006	0,057	0,107	0,915
HHSTAT3	-0,061	0,072	-0,841	0,4
FIRMPRI3	-0,081	0,085	-0,952	0,341
CONTSTRO1	-0,267	0,226	-1,182	0,238
TENUEXPE	-0,006	0,005	-1,411	0,159
CONTSTRO7	-0,549	0,312	-1,757	0,079
Occtype3	-0,099	0,042	-2,34	0,019
FIRMPRI2	-0,218	0,063	-3,466	0,001
Occtype1	-0,155	0,034	-4,549	< 0,0001
Occtype2	-0,294	0,046	-6,354	< 0,0001
R²	0,277	-	-	-
Observations	1134	-	F	Pr > F
Fisher	-	-	20,259	< 0,0001

Note 4: Les résultats proviennent de l'estimation d'un modèle M.C.O. Le logiciel utilisé est XLSTAT. La variable expliquée est le logarithme du salaire horaire brut des individus, les variables explicatives sont pour chaque individus : l'âge, le fait qu'il ait ou non des enfants, un contrat à durée déterminée ou non, la position de son emploi dans la hiérarchie de l'entreprise, la durée de ses études, le pourcentage de femmes dans l'entreprise, si l'entreprise relève du secteur privé ou public, le sexe de l'individu, son statut marital, son expérience dans le milieu du travail.

3.2.3. Régression sur le groupe femmes

Tableau 8: Régression sur les femmes

Source	Valeur	Ecart-type	t	Pr > t
Constante	1,741	0,184	9,462	< 0,0001
EDUCATYR	0,059	0,006	9,318	< 0,0001
firmfema40-60	0,113	0,038	2,988	0,003
CONTSTRO7	0,577	0,207	2,786	0,005
TENUEXPE	0,01	0,004	2,744	0,006
HHSTAT1	0,058	0,03	1,97	0,049
firmfema20-40	0,081	0,042	1,915	0,056
HRSHISTO	0,062	0,032	1,894	0,058
CONTR7	0,064	0,042	1,505	0,133
firmfema60-80	0,063	0,042	1,499	0,134
HHSTAT2	0,247	0,174	1,421	0,156
AGE	0,005	0,004	1,411	0,159
FIRMPRI2	0,063	0,049	1,293	0,196
FIRMPRI1	0,033	0,026	1,27	0,204
firmfema0-20	0,046	0,044	1,035	0,301
FIRMPRI3	0,021	0,047	0,452	0,651
chld	0,011	0,032	0,342	0,733
CONTSTRO1	-0,073	0,136	-0,535	0,593
HHSTAT3	-0,046	0,053	-0,861	0,39
Occtype3	-0,096	0,055	-1,739	0,082
CONTSTRO8	-0,734	0,41	-1,788	0,074
Occtype2	-0,275	0,042	-6,573	< 0,0001
Occtype1	-0,255	0,037	-6,915	< 0,0001
R ²	0,277	-	-	-
Observations	1134	-	F	Pr > F
Fisher	-	-	20,259	<0,0001

Note 5: Les résultats proviennent de l'estimation d'un modèle M.C.O. Le logiciel utilisé est XLSTAT. La variable expliquée est le logarithme du salaire horaire brut des individus, les variables explicatives sont pour chaque individus : l'âge, le fait qu'il ait ou non des enfants, un contrat à durée déterminée ou non, la position de son emploi dans la hiérarchie de l'entreprise, la durée de ses études, le pourcentage de femmes dans l'entreprise, si l'entreprise relève du secteur privé ou public, le sexe de l'individu, son statut marital, son expérience dans le milieu du travail.

3.2.4. Méthode d'Oaxaca-Blinder

Les estimations hommes et femmes ayant été réalisées séparément, les éventuelles différences entre les coefficients vont être décomposées suivant la méthode d'Oaxaca-Blinder.

Variables	Coefficients Hommes	Coefficients Femmes	Déférence des coefficients	Moyenne Hommes	Moyenne Femmes	Déférence des moyenne	Discrimination F ⁹	Discrimination H ¹⁰
Constante	1,675	1,741	-0,066				-0,066	-0,066
AGE	0,022	0,005	0,017	34,074	33,885	0,189	0,570	0,570
chld	0,052	0,011	0,041	0,459	0,490	-0,030	0,018	0,018
CONTR7	0,006	0,064	-0,058	0,936	0,907	0,028	-0,052	-0,052
CONTSTRO1	-0,267	-0,073	-0,194	0,990	0,985	0,005	-0,193	-0,193
CONTSTRO7	-0,549	0,577	-1,126	0,004	0,006	-0,002	-0,006	-0,006
CONTSTRO8	0,383	-0,734	1,117	0,002	0,001	0,001	0,001	0,001
CONTSTRO9	0,000	0,000	0,000	0,004	0,008	-0,004	0,000	0,000
EDUCATYR	0,060	0,059	0,000	13,817	13,836	-0,019	0,004	0,004
firmfema0-20	0,025	0,046	-0,021	0,389	0,150	0,239	0,003	0,003
firmfema20-40	0,097	0,081	0,016	0,264	0,190	0,073	0,010	0,010
firmfema40-60	0,085	0,113	-0,029	0,252	0,330	-0,078	-0,016	-0,016
firmfema60-80	0,028	0,063	-0,035	0,072	0,176	-0,104	-0,009	-0,009
firmfema80-100	0,000	0,000	0,000	0,023	0,154	-0,131	0,000	0,000
FIRMPRI1	0,008	0,033	-0,026	0,453	0,485	-0,031	-0,013	-0,013
FIRMPRI2	-0,218	0,063	-0,281	0,052	0,074	-0,022	-0,016	-0,016
FIRMPRI3	-0,081	0,021	-0,103	0,027	0,081	-0,053	-0,004	-0,004
FIRMPRIvide	0,000	0,000	0,000	0,467	0,361	0,106	0,000	0,000
HHSTAT1	0,083	0,058	0,025	0,536	0,550	-0,014	0,012	0,012
HHSTAT2	0,000	0,247	-0,247	0,000	0,005	-0,005	-0,001	-0,001
HHSTAT3	-0,061	-0,046	-0,015	0,046	0,077	-0,031	0,001	0,001
HHSTAT4	0,000	0,000	0,000	0,418	0,368	0,050	0,000	0,000
HRSHISTO	0,063	0,062	0,001	0,974	0,811	0,164	0,011	0,011
Occtype1	-0,155	-0,255	0,100	0,426	0,544	-0,118	0,073	0,073
Occtype2	-0,294	-0,275	-0,019	0,147	0,245	-0,098	0,024	0,024
Occtype3	-0,099	-0,096	-0,004	0,161	0,076	0,086	-0,009	-0,009
Occtype4	0,000	0,000	0,000	0,265	0,135	0,131	0,000	0,000
TENUEXPE	-0,006	0,010	-0,016	11,880	11,615	0,265	-0,190	-0,190
Ln (WAGE) moyen				2,960798628	2,80677507			
	$\overline{\ln W_m} - \overline{\ln W_f} = 0,154023557$						0,154023557	0,15402356

Note 6: Les résultats proviennent de l'estimation de la décomposition d'Oaxaca-Blinder. La variable expliquée est le logarithme du salaire horaire brut des individus, les variables explicatives sont pour chaque individu: l'âge, le fait qu'il ait ou non des enfants, un contrat à durée déterminée ou non, la position de son emploi dans la hiérarchie de l'entreprise, la durée de ses études, le pourcentage de femmes dans l'entreprise, si l'entreprise relève du secteur privé ou public, le sexe de l'individu, son statut marital, son expérience dans le milieu du travail.

⁹L'équation suivante mesure la *discrimination des femmes*:

$$\overline{\ln W_m} - \overline{\ln W_f} = \widehat{\alpha}_m - \widehat{\alpha}_f + (\widehat{\beta}_m - \widehat{\beta}_f) \cdot \bar{X}_f + \widehat{\beta}_m (\bar{X}_m - \bar{X}_f)$$

¹⁰L'équation suivante mesure la *discrimination des hommes*:

$$\overline{\ln W_m} - \overline{\ln W_f} = \widehat{\alpha}_m - \widehat{\alpha}_f + (\widehat{\beta}_m - \widehat{\beta}_f) \cdot \bar{X}_m + \widehat{\beta}_f (\bar{X}_m - \bar{X}_f)$$

Interprétation des résultats:

La dissemblance des deux régressions séparées, tant sur les hommes que sur les femmes, montre que l'effet d'une variable peut être différent selon le sexe et justifie donc l'utilisation de la méthode de la décomposition d'Oaxaca-Blinder.

Cette méthode permet de calculer les différences entre les moyennes des attributs des hommes et celles des femmes dans l'estimation du différentiel salarial moyen.

Dans notre étude, elle nous permet de constater que 15,4 % de la différence totale entre le salaire des hommes et des femmes ne sont pas expliqués par des différences dans la productivité moyenne, mais peuvent être attribués à une discrimination salariale envers les femmes.

Certaines variables indiquent une discrimination ponctuelle envers les hommes (différence des coefficients négatifs). Néanmoins, c'est la différence totale qui permet de conclure à l'encontre de quel sexe la discrimination salariale se produit, et précisément à l'encontre des femmes.

Conclusion

L'objet de notre étude était de comprendre: "*D'où viennent les inégalités de salaires entre hommes et femmes en Belgique?*".

L'étude économétrique mise en œuvre via la méthode des moindres carrés ordinaires nous permet d'affirmer que la discrimination salariale envers les femmes est de l'ordre de 10%, ces chiffres étant corroborés par les deux régressions linéaires consécutives. La décomposition d'Oaxaca-Blinder nous propose des chiffres plus élevés encore, puisque nous relevons par cette méthode, une discrimination à l'encontre des femmes de 15,4 %. En conclusion, le sexe apparaît bien être un facteur déterminant dans les écarts salariaux.

Toutefois, cette étude comporte des limites:

- ***récolte des données***

Les données ont été récoltées à l'aide d'un questionnaire en ligne uniquement. Cela suppose que les personnes qui ont participé à l'enquête ont accès à internet. Or, bon nombre de travailleurs n'y ont pas nécessairement accès.

Par ailleurs, les personnes qui remplissent le questionnaire sont des personnes qui consultent le site internet national du WageIndicator. Ces personnes se rendent sur le site pour obtenir de l'information sur les salaires. Il s'agit généralement de salariés qui préparent leur évaluation de performance annuelle ou leur négociation de salaire. Le site est également consulté par des étudiants, des employeurs de petites et moyennes entreprises lorsqu'ils souhaitent recruter du personnel ou négocier le salaire de leurs employés,...

Il est important de noter qu'en retour d'information gratuite sur les salaires, toutes les personnes visitant le site sont invitées à remplir le questionnaire en ligne, opération qui prend 10 à 20 minutes. C'est de cette façon que les données sont récoltées. Ce questionnaire comporte des questions sur, notamment, les rémunérations, les conditions d'emploi, le type de contrat et formation, le niveau d'éducation, le type d'industrie, les caractéristiques familiales (marié(e) ou pas, nombre d'enfants,...), le sexe, l'âge,....

- ***biais de sélection***

Les personnes sélectionnées pour l'étude sont uniquement des salariés. Dès lors, les personnes qui ne travaillent pas ne sont pas prises en compte. Or, on peut imaginer qu'une femme ait décidé de ne pas travailler car le coût d'opportunité d'aller travailler n'est pas suffisant. Concrètement, si une femme perçoit une rémunération mensuelle de 1.250€ et que le coût de crèche où elle dépose son enfant 5 jours par semaine est de 900€ par mois, la probabilité qu'elle ne travaille pas est élevée. On peut donc dire que l'échantillon ne représente pas, dans ce cas, la population.

La méthode de Heckman permet toutefois de détecter la présence d'un éventuel biais de sélection et de corriger ce biais s'il est présent. Elle consiste à estimer la probabilité qu'à un individu d'être sur le marché du travail et à utiliser cette probabilité comme variable explicative lors de la régression effectuée sur l'échantillon complet, c'est-à-dire l'échantillon qui comprend également les inactifs et les chômeurs).

- ***limites de la méthode***

Les méthodes d'analyse économétriques vues au cours ne sont pas suffisantes dans le cadre de notre étude, en effet, les hypothèses de Gauss-Markov suivantes sont toutes les deux violées:

- H4: absence d'autocorrélation entre les termes d'erreur et homoscédasticité des erreurs, c'est-à-dire variance constante sur l'échantillon;
- H6: terme d'erreur suit une loi normale d'espérance nulle et de variance σ^2 .

La piste des Moindres Carrés Généraux en ce qui concerne l'hétéroscédasticité devrait être une solution plausible, ainsi qu'un test de Hausman. Toutefois, les contraintes de temps et les limites de notre connaissance ne nous permettent pas de les appliquer.

Annexes

Annexe 1: Etudes économétriques menées en France et aux Etats-Unis.....	29
Annexe 2: Source des données.....	32
Annexe 3: Divers graphiques	33
Annexe 4: Test de normalité de Jarque-Bera appliqué au modèle (1)	36
Annexe 5: Tableau de résultats de la régression sur les paramètres significatifs et tableau des variables significatives au seuil de 5%	37
Annexe 6: Test de significativité global des coefficients de Fisher appliqué au modèle (1).....	38
Annexe 7: Matrice de corrélation	39

Annexe 1

Les tableaux¹¹ ci-après présentent quelques études économétriques qui ont été menées entre 1973 et 2000 en France et aux Etats-Unis:

	Ronald Oaxaca (<i>International Economic</i> , vol 14, n°3, 1973)	Jacob Mincer, Solomon Polacheck (<i>The Journal of Political Economy</i> , vol 82, n° 2, 1974)	Bernard Thiry (<i>Annales de l'INSEE</i> , n° 58, 1985)	Michel Glaude (<i>Données sociales</i> , 1987)
Source	Survey of economic opportunity, 1967, Etats-Unis	National longitudinal survey of work experience, 1967 et Survey of economic opportunity, 1966, Etats-Unis	Enquête structure des salaires, 1978, France	Enquête emploi, 1985, France
Champ	Salariés blancs ou noirs âgés de plus de 16 ans vivant en zone urbaine	Femmes mariées, hommes mariés et femmes célibataires âgés de 30 à 44 ans	Seulement une partie de la population couverte par l'enquête: les travailleurs manuels à plein temps des entreprises de plus de 10 salariés dans l'industrie	Tous les salariés à temps plein
Grandeur estimée	Salaire horaire	Salaire horaire	Salaire horaire	Salaire mensuel
Variables explicatives retenues dans l'équation de salaire	<u>Modèle 1</u> : expérience, éducation, santé, temps partiel, migration, statut marital, taille de la ville, région <u>Modèle 2</u> : Modèle 1 + occupation (proche des catégories socio-professionnelle en 10 postes) et secteur (16 postes)	Expérience, années de scolarité	<u>Modèle 1</u> : âge, ancienneté, proportion d'ouvriers ayant telle ancienneté (8 postes), proportion d'étrangers, travail en équipe (3 postes), proportion d'ouvriers payés au rendement, proportion d'ouvriers mensualisés <u>Modèle 2</u> : modèle 1 + qualification (qualifié, semi-qualifié, non qualifié) et secteur d'activité en 44 postes	Catégorie socio-professionnelle fonction (37 postes), diplôme (16 postes), ancienneté, durée du travail, secteur, statut de l'entreprise, taille de l'entreprise, expérience, type de ménage, activité de l'épouse, nationalité, région, catégorie de commune

¹¹ DE CURRAIZE Y. et HUGOUNENG R. (2004), pp. 206-208

	Ronald Oaxaca	Jacob Mincer, Solomon Polachek	Bernard Thiry	Michel Glaude
Résultats	<p>En 1967, les hommes blancs ont un salaire moyen de 54% supérieur aux femmes blanches et les hommes noirs de 49% supérieur aux femmes noires. En utilisant le <u>modèle 1</u>, 77% (blancs) et 93% (noirs) du différentiel reste inexpliqué et peut être attribué à la discrimination. Le <u>modèle 2</u>, qui ne tient pas compte de la discrimination à l'embauche, donne un écart inexpliqué et attribuable à la discrimination de 58% (blancs) et 55% (noirs)</p>	<p>Le salaire des hommes mariés est de 52% supérieur à celui des femmes mariées et de 16% supérieur à celui des femmes célibataires. La part de l'écart total inexpliquée par le modèle et attribuable à la discrimination est comprise entre 55% (coefficients des hommes) et 58% (coefficients des femmes) pour la comparaison hommes mariés/femmes mariées, et entre 60% (coefficients des femmes) et 93% (coefficients des hommes) pour la comparaison hommes mariés/femmes célibataires</p>	<p>L'écart total en terme de salaire horaire est de 30% par rapport au salaire masculin. L'écart inexpliqué se réduit à 23% avec le <u>modèle 1</u>, soit 76% de l'écart total et à 12%-15% avec le <u>modèle 2</u>, soit 40% à 50% de l'écart total</p>	<p>En 1985, les hommes ont un salaire moyen de 25% supérieur à celui des femmes. L'écart inexpliqué par le modèle est de + 14,5% en faveur des hommes, soit 58% de l'écart total</p>

					Catherine Sofer (<i>Economie et Prévision</i> , n°92-93, 1990)	Jean-Louis Lhéritier (<i>Economie et Statistique</i> , n° 257, 1992)	Alain Bayet (<i>Données sociales</i> , 1996)	Dominique Meurs, Sophie Ponthieux (<i>Economie et Statistique</i> , n° 337-338, 2000)
Source	Enquête FQP ¹² , 1977, France		Enquête sur la structure des salaires, 1986, France		Enquête sur la structure des salaires, 1992, France		Enquête jeunes et carrière, 1997, France	
Champ	Salariés de tout âge n'ayant pas changé d'employeur avant 1972 ou entrés dans la vie active entre 1972 et 1977		Salariés à temps plein de l'industrie et des services âgés de 18 à 59 ans		Salariés à temps plein de l'ensemble du secteur marchand		Salariés hors enseignants âgés de moins de 45 ans	
Grandeur estimée	Salaire annuel		Salaire mensuel		Salaire mensuel		Salaire mensuel	
Variables explicatives retenues dans l'équation de salaire	Éducation (en années), expérience (durée réelle), durée totale de chômage, durée totale d'inactivité, temps passé hors marché du travail (chômage + inactivité), ancienneté, taux de féminisation du dernier emploi occupé		Catégorie socio-professionnelle, qualification (mais ce n'est pas le diplôme), expérience, ancienneté, nationalité, système de rémunération, secteur, taille de l'entreprise, certaines caractéristiques de l'employeur (existence de convention collective, d'accord d'entreprise, part d'ouvriers, d'étrangers et de femmes dans l'effectif de l'entreprise, existence de travail en équipe ou de nuit), département		Catégorie socio-professionnelle, âge, diplôme, ancienneté, régime horaire, secteur, taille de l'entreprise, région, caractéristiques de l'employeur concernant sa politique salariale et sa position sur le marché (théorie de la rente)		Le diplôme en 8 postes dont 3 diplômes du supérieur (licence, grande école), expérience par composantes (expérience effective, ancienneté, durée de chômage, durée d'inactivité, autre durée hors emploi), enfants, couple, Paris, catégorie socio-professionnelle en 6 postes, contrat en 4 postes, durée hebdomadaire de travail en 5 postes dont la raison du temps partiel, secteur taille en 11 postes, conditions de travail, nationalité	
Résultats	L'écart total est de 18,6% et l'écart inexplicable est compris entre 9,3% et 20,3% , soit au minimum 50% de l'écart total. En introduisant le taux de féminisation, l'écart inexplicable est compris entre 0,7% et 17,8% , soit de 3% à 95% de l'écart total		L'écart inexplicable (la discrimination salariale) est de + 12,5% pour les hommes dans l'industrie, de + 11,3% dans les services		L'écart inexplicable par le modèle est de + 14% en faveur des hommes		L'écart total de 27% en faveur des hommes tombe à 16% à durée hebdomadaire du travail égale, soit 59% de l'écart total. En contrôlant par l'ensemble des variables du modèle, l'écart inexplicable est de 5,4% soit 20% de l'écart total.	

¹² Enquête FQP signifie Enquête sur la Formation et la Qualification Professionnelle

Annexe 2

Source des données

La recherche de la base de données nous permettant de pouvoir travailler a été fastidieuse. En effet, le voile de pudeur attaché à la question des salaires est très difficile à lever dans notre pays.

Nous avons dressé les sources contactées avec la chronologie suivante:

- Ministère Fédéral de l'Emploi et du Travail, Direction de l'Egalité des Chances (Madame Marie-Paule Urbain);
- Direction Générale des Statistiques du Gouvernement fédéral (Monsieur Pieter Vermeulen);
- Eurostat Unit F2, Head of section « Earning labour costs, PEEI for the labour market” statistics (Monsieur Didier Dupré).

La longueur des procédures ainsi que la lourdeur administrative pour introduire une demande dont nous ne savons si elles seront couronnées du succès escompté sont décourageantes.

Notre démarche ne fût pas une réussite auprès des institutions officielles, nous avons dès lors décidé de donner une autre direction à nos recherches.

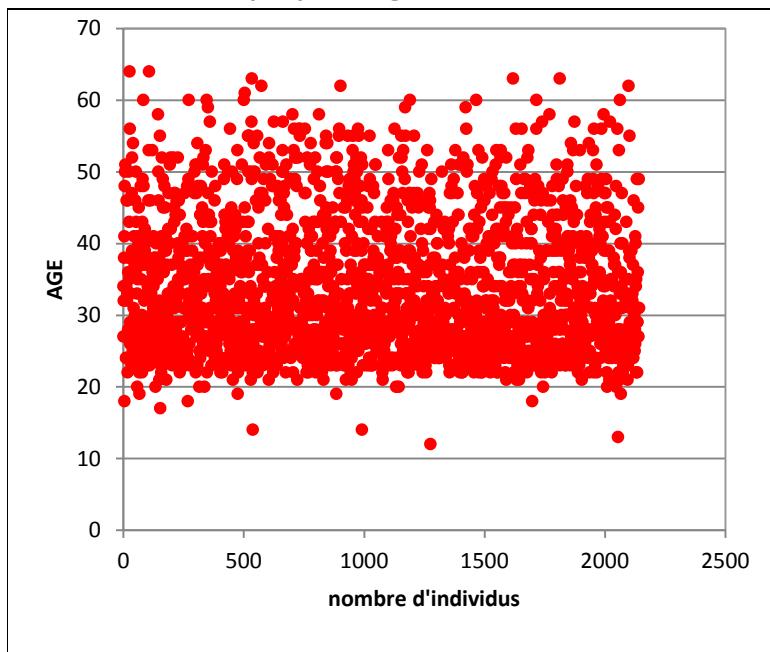
Dans ce contexte, nous avons fait le tour de différentes universités du pays en vain en terminant par une prise de contact avec l'Université d'Anvers en la personne du Professeur Dimitri Mortelmans, PhD responsable du « Research Centre for Longitudinal and Life Course Studies (CELLO) ». Celui-ci nous a alors suggéré de consulter l'adresse suivante www.PSHBE.be ainsi que de chercher d'autres études qui auraient été réalisées en la matière.

En parallèle avec les premiers contacts, nous avions retenu dans nos favoris un site belge www.votresalaire.be avec comme interlocutrice Madame Léontine Bijleveld. Celle-ci nous a proposé de nous mettre en relation avec le Professeur Kea Tijdens de l'Université d'Amsterdam.

Nous avions alors une piste sérieuse et ,après quelques e-mails échangés ainsi qu'une clause de confidentialité et de collaboration signée, nous disposions enfin après tous ces écueils de données relatives à la Belgique sur les salaires.

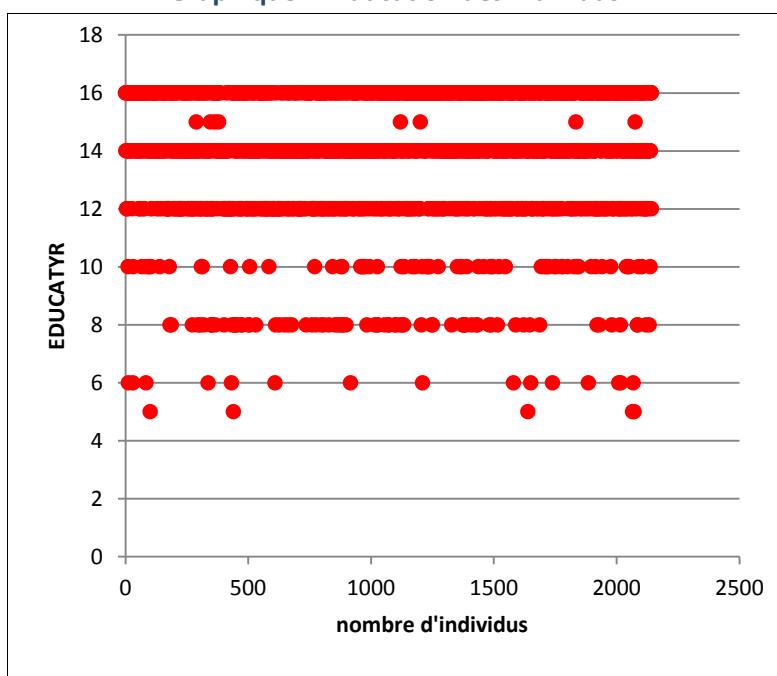
Annexe 3

Graphique 3: Age des Individus



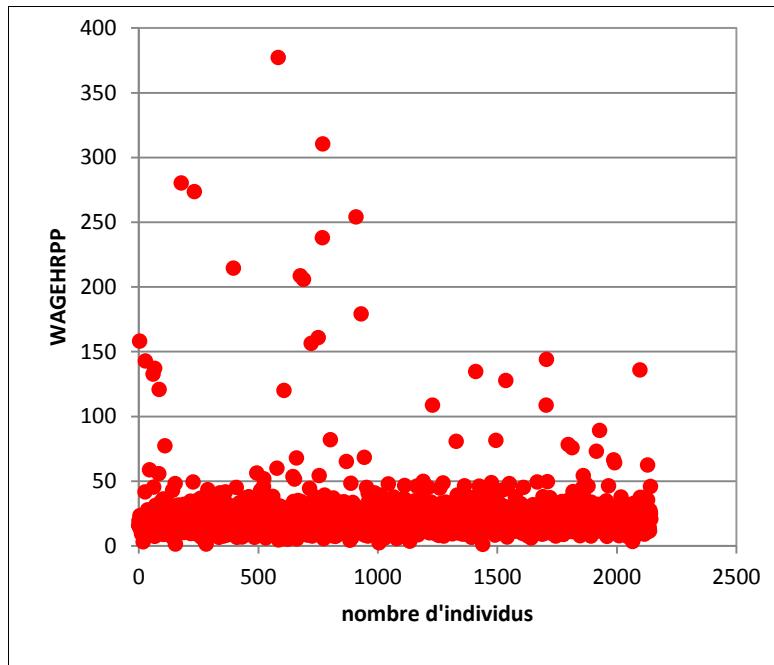
Source: Tijdens et Osse (2010)

Graphique 4: Education des individus



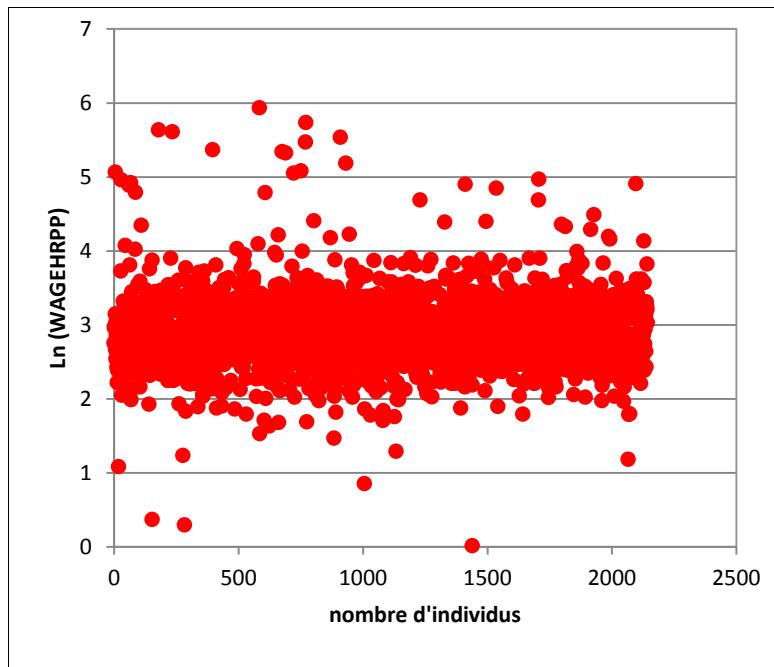
Source: Tijdens et Osse (2010)

Graphique 5: Salaire horaire brut des individus



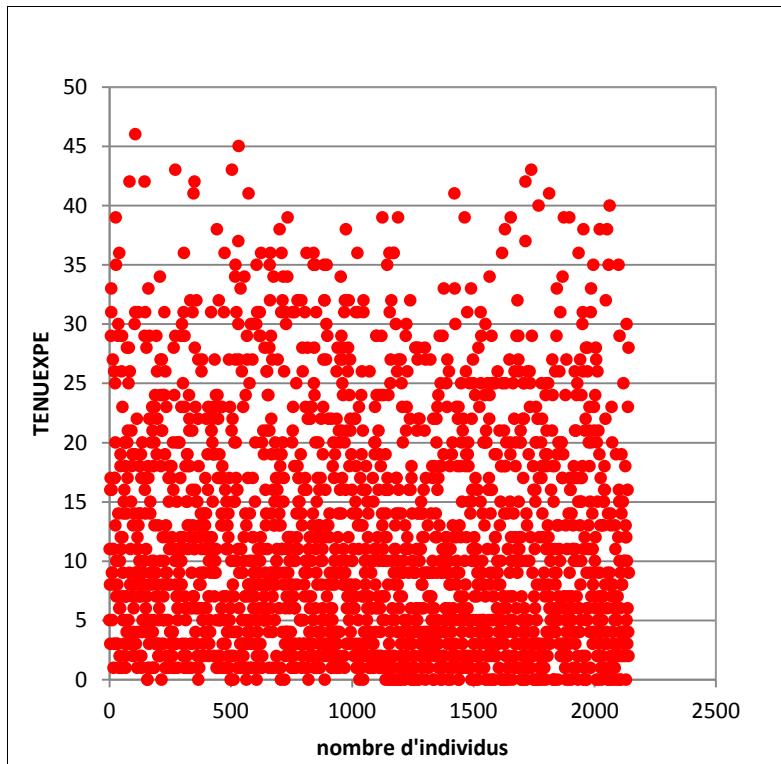
Source: Tijdens et Osse (2010)

Graphique 6: Logarithme népérien du salaire horaire brut des individus



Source: Tijdens et Osse (2010)

Graphique 7: Expérience des individus



Source: *Tijdens et Osse (2010)*

Annexe 4

Test de normalité de Jarque-Bera appliqué au modèle (1)

Pour ce test, il est nécessaire dans un premier temps de collecter les résidus de l'estimation. Le test de Jarque-Bera est un test d'hypothèse qui cherche à déterminer si les données suivent une loi normale ou non. En l'occurrence ici, nous allons tester la normalité des résidus afin de vérifier l'hypothèse H6.

Il consiste à tester l'hypothèse nulle :

$$\begin{cases} H_0: \text{la série de données suit une loi normale} \\ H_1: \text{la série de données ne suit pas une loi normale} \end{cases}$$

Le test de Jarque-Bera est basé sur la statistique suivante :

$$JB = \frac{n - k}{6} \left[S^2 + \frac{(K - 3)^2}{4} \right]$$

Sous l'hypothèse nulle, cette statistique suit asymptotiquement une loi du χ^2 à 2 degrés de liberté.

Avec n le nombre d'observations, k le nombre de variables explicatives du modèle, S le coefficient d'asymétrie (skewness, moment d'ordre 3 d'une variable centrée-réduite), et K, le kurtosis (moment d'ordre 4 d'une variable centrée réduite).

Ainsi, si les données sont normales, la statistique s'approche de zéro et diverge sinon.

Test de Jarque-Bera: sur les résidus

JB (valeur observée)	12638,956
JB (valeur critique)	5,991
ddl	2
p-value unilatérale	< 0,0001
Alpha	0,05

Conclusion:

Au seuil de signification de 5% (il en va de même pour un seuil de 10% !)

On peut rejeter l'hypothèse nulle selon laquelle l'échantillon suit une loi normale.

Autrement dit, la non-normalité est significative.

Vu les résultats présentés, on rejette l'hypothèse H_0 de normalité des résidus. Il est devrait donc être nécessaire de mettre en œuvre une autre procédure d'estimation.

Mais celles-ci n'ont pas été évoquées en cours, c'est pourquoi nous considérerons malgré tout que l'hypothèse H6 est vérifiée.

Annexe 5

Tableau 6: Résultats de la régression sur les paramètres significatifs

Paramètre	Valeur	Ecart-type	t de Student	Pr > t
Constante	1,638	0,085	19,206	< 0,0001
AGE	0,015	0,001	13,915	< 0,0001
EDUCATYR	0,059	0,004	13,445	< 0,0001
firmfema20-40	0,062	0,023	2,666	0,008
firmfema40-60	0,072	0,021	3,341	0,001
GENDER	-0,101	0,019	-5,193	< 0,0001
HHSTAT1	0,093	0,019	4,783	< 0,0001
HRSHISTO	0,060	0,032	1,908	0,057
Occtype1	-0,210	0,025	-8,457	< 0,0001
Occtype2	-0,277	0,030	-9,119	< 0,0001
Occtype3	-0,092	0,033	-2,808	0,005

Note 7: tableau repris en page 21

Tableau 9 : Variables significatives au seuil de 5%

Source	Valeur	Ecart-type	t	Pr > t
Constante	1,692	0,156	10,861	< 0,0001
AGE	0,012	0,003	3,877	0,000
EDUCATYR	0,061	0,005	13,042	< 0,0001
firmfema20-40	0,100	0,038	2,597	0,009
firmfema40-60	0,110	0,037	2,984	0,003
GENDER	-0,096	0,021	-4,638	< 0,0001
HHSTAT1	0,073	0,023	3,120	0,002
HRSHISTO	0,075	0,032	2,318	0,021
Occtype1	-0,203	0,025	-8,132	< 0,0001
Occtype2	-0,272	0,031	-8,836	< 0,0001
Occtype3	-0,095	0,033	-2,861	0,004

Annexe 6

Test de significativité global des coefficients de Fisher appliqué au modèle (1) présenté dans le tableau (5) en page 19

On teste $\begin{cases} H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \dots \beta_j = \dots \beta_n = 0 \\ \text{Contre } H_a: \exists_j, \beta_j \neq 0, \text{ il existe au moins un } j \text{ tel que } \beta_j \neq 0 \end{cases}$

Pour cela, on calcule la F-stat :

$$F = \frac{SSR_C - SSR_{NC}}{SSR_{NC}} * \frac{(N - K - 1)}{J}$$

Ou

$$F = \frac{R_{NC}^2 - R_C^2}{1 - R_{NC}^2} * \frac{(N - K - 1)}{J}$$

Avec SSR_{NC} la somme des carrés des résidus non contraints, SSR_C , la somme des résidus contraints (sans les J variables), K représentant le nombre de variables et N le nombre d'observations.

Sous l'hypothèse nulle (i.e. si l'hypothèse nulle est vraie) cette statistique suit une distribution connue.

$$F(J, N - K - 1)$$

Où $F(J, N - K - 1)$ est une distribution de Fisher, N étant le nombre d'observations dans l'échantillon, J le nombre de variables du modèle contraint (sans les variables) et K le nombre de régresseurs du modèle non contraint (i.e. avec toutes les variables).

La valeur de la F-stat est présentée dans le tableau 3 précédent. Le logiciel XLSTAT présente également la p-value correspondante.

La p-value fourni la probabilité de rejeter l'hypothèse nulle à tort. Pour procéder au test, il faut définir une règle de décision.

Si la p-valeur est inférieure à 5%, on décide de **rejeter l'hypothèse nulle**.

A l'inverse, **on ne pourra rejeter l'hypothèse nulle** si la p-valeur est supérieure à 5%.

Annexe 7

Matrice de Corrélation

Variables	AGE	chld	CONTR7	CONSTRO1	CONSTRO7	CONSTRO8	CONSTRO9	EDUCATYR	firmfema0-20	firmfema20-40	firmfema40-60	firmfema60-80	firmfema80-100	FIRMPRI1	FIRMPRI2	FIRMPRI3	FIRMPRIvide	GENDER	HHSTAT1	HHSTAT2	HHSTAT3	HHSTAT4	HRSHISTO	Occtype1	Occtype2	Occtype3	Occtype4	TENUEXPE	Ln (WAGEHRPP)
AGE	1,000	0,564	0,168	0,043	-0,090	-0,010	0,029	-0,181	0,001	0,030	-0,043	-0,013	0,038	0,058	0,086	-0,011	-0,096	-0,008	0,332	0,081	0,243	-0,465	-0,119	-0,070	-0,201	0,147	0,165	0,936	0,326
chld	0,564	1,000	0,129	0,037	-0,068	-0,035	0,029	-0,160	0,035	0,001	-0,066	0,002	0,049	0,053	0,053	-0,014	-0,073	0,030	0,482	0,051	0,173	-0,580	-0,145	-0,041	-0,183	0,062	0,181	0,544	0,215
CONTR7	0,168	0,129	1,000	0,141	-0,148	-0,035	-0,047	0,019	0,016	0,017	0,015	-0,018	-0,054	0,088	-0,074	-0,054	-0,028	-0,053	0,122	-0,022	-0,006	-0,119	0,060	-0,025	-0,127	0,077	0,093	0,157	0,143
CONSTRO1	0,043	0,037	0,141	1,000	-0,648	-0,338	-0,677	-0,017	0,030	0,000	-0,014	-0,011	-0,012	0,010	-0,007	0,026	-0,019	-0,024	0,061	-0,083	0,010	-0,059	0,131	-0,013	-0,021	0,041	0,003	0,035	-0,012
CONSTRO7	-0,090	-0,068	-0,148	-0,648	1,000	-0,003	-0,005	0,042	-0,030	-0,008	0,041	0,013	-0,022	-0,028	-0,019	-0,017	0,045	0,011	-0,065	-0,003	-0,018	0,075	-0,169	0,009	0,031	-0,027	-0,020	-0,060	0,000
CONSTRO8	-0,010	-0,035	-0,035	-0,338	-0,003	1,000	-0,003	0,015	0,005	0,009	-0,024	0,024	-0,011	-0,010	-0,010	-0,009	0,019	-0,010	-0,041	-0,002	-0,009	0,046	0,013	-0,036	0,076	-0,014	-0,019	-0,044	-0,002
CONSTRO9	0,029	0,029	-0,047	-0,677	-0,005	-0,003	1,000	-0,023	-0,018	0,004	-0,006	-0,009	0,045	0,017	0,032	-0,018	-0,025	0,030	-0,006	0,126	0,007	-0,010	-0,036	0,028	-0,028	0,024	0,028	0,019	
EDUCATYR	-0,181	-0,160	0,019	-0,017	0,042	0,015	-0,023	1,000	-0,076	0,045	0,092	-0,013	-0,081	-0,064	0,013	0,025	0,047	0,003	-0,042	-0,032	-0,106	0,098	0,122	-0,094	0,044	0,055	0,029	-0,280	0,216
firmfema0-20	0,001	0,035	0,016	0,030	-0,030	0,005	-0,018	-0,076	1,000	-0,338	-0,394	-0,229	-0,188	0,052	-0,095	-0,123	0,050	-0,268	-0,005	-0,008	-0,007	0,010	0,092	-0,042	-0,053	0,065	0,051	0,037	-0,007
firmfema20-40	0,030	0,001	0,017	0,000	-0,008	0,009	0,004	0,045	-0,338	1,000	-0,348	-0,202	-0,166	0,004	-0,022	-0,059	0,033	-0,086	0,005	-0,003	0,025	-0,017	0,046	-0,072	0,021	0,087	-0,002	0,016	0,063
firmfema40-60	-0,043	-0,066	0,015	-0,014	0,041	-0,024	-0,006	0,092	-0,394	-0,348	1,000	-0,236	-0,193	-0,012	0,057	-0,007	-0,012	0,085	-0,025	0,012	-0,044	0,046	0,012	0,045	0,029	-0,043	-0,049	-0,059	0,038
firmfema60-80	-0,013	0,002	-0,018	-0,011	0,013	0,024	-0,009	-0,013	-0,229	-0,202	-0,236	1,000	-0,113	-0,026	0,040	0,124	-0,049	0,159	-0,001	-0,018	0,021	-0,007	-0,063	0,022	0,024	-0,072	0,007	-0,020	-0,041
firmfema80-100	0,038	0,049	-0,054	-0,012	-0,022	-0,011	0,045	-0,081	-0,188	-0,166	-0,193	-0,113	1,000	-0,039	0,046	0,154	-0,053	0,236	0,043	0,020	0,022	-0,057	-0,163	0,077	-0,022	-0,082	-0,008	0,036	-0,098
FIRMPRI1	0,058	0,053	0,088	0,010	-0,028	-0,010	0,017	-0,064	0,052	0,004	-0,012	-0,026	-0,039	1,000	-0,242	-0,221	-0,793	0,032	0,037	-0,007	0,026	-0,050	0,000	-0,001	-0,011	0,057	-0,034	0,077	0,037
FIRMPRI2	0,086	0,053	-0,074	-0,007	-0,019	-0,010	0,032	0,013	-0,095	-0,022	0,057	0,040	0,046	-0,242	1,000	-0,061	-0,218	0,047	0,040	0,027	0,016	-0,051	-0,008	0,056	-0,029	-0,078	0,022	0,074	-0,006
FIRMPRI3	-0,011	-0,014	-0,054	0,026	-0,017	-0,009	-0,018	0,025	-0,123	-0,059	-0,007	0,124	0,154	-0,221	-0,061	1,000	-0,200	0,117	0,011	0,032	-0,042	0,006	-0,051	0,028	0,006	-0,068	0,015	-0,016	-0,038
FIRMPRIvide	-0,096	-0,073	-0,028	-0,019	0,045	0,019	-0,025	0,047	0,050	0,033	-0,012	-0,049	-0,053	-0,793	-0,218	-0,200	1,000	-0,109	-0,062	-0,021	-0,015	0,073	0,027	-0,039	0,023	0,011	0,017	-0,107	-0,017
GENDER	-0,008	0,030	-0,053	-0,024	0,011	-0,010	0,030	0,003	-0,268	-0,086	0,085	0,159	0,236	0,032	0,047	0,117	-0,109	1,000	0,015	0,051	0,065	-0,052	-0,271	0,120	0,123	-0,130	-0,163	-0,013	-0,156
HHSTAT1	0,332	0,482	0,122	0,061	-0,065	-0,041	-0,006	-0,042	-0,005	0,005	-0,025	-0,001	0,043	0,037	0,040	0,011	-0,062	0,015	1,000	-0,053	-0,275	-0,880	-0,058	-0,035	-0,160	0,105	0,115	0,328	0,211
HHSTAT2	0,081	0,051	-0,022	-0,083	-0,003	-0,002	0,126	-0,032	-0,008	-0,003	0,012	-0,018	0,020	-0,007	0,027	0,032	-0,021	0,051	-0,053	1,000	-0,012	-0,039	-0,047	-0,008	0,025	-0,018	-0,001	0,075	0,028
HHSTAT3	0,243	0,173	-0,006	0,010	-0,018	-0,009	0,007	-0,106	-0,007	0,025	-0,044	0,021	0,022	0,026	0,016	-0,042	-0,015	0,065	-0,275	-0,012	1,000	-0,205	-0,011	0,024	-0,025	-0,016	0,008	0,229	-0,007
HHSTAT4	-0,465	-0,580	-0,119	-0,059	0,075	0,046	-0,010	0,098	0,010	-0,017	0,046	-0,007	-0,057	-0,050	-0,051	0,006	0,073	-0,052	-0,880	-0,039	-0,205	1,000	0,069	0,025	0,173	-0,098	-0,121	-0,453	-0,214
HRSHISTO	-0,119	-0,145	0,060	0,131	-0,169	0,013	-0,036	0,122	0,092	0,046	0,012	-0,063	-0,163	0,000	-0,008	-0,051	0,027	-0,271	-0,058	-0,047	-0,011	0,069	1,000	-0,096	-0,013	0,064	0,080	-0,132	0,085
Occtype1	-0,070	-0,041	-0,025	-0,013	0,009	-0,036	0,028	-0,094	-0,042	-0,072	0,045	0,022	0,077	-0,001	0,056	0,028	-0,039	0,120	-0,035	-0,008	0,024	0,025	-0,096	1,000	-0,472	-0,357	-0,488	-0,066	-0,145
Occtype2	-0,201	-0,183	-0,127	-0,021	0,031	0,076	-0,037	0,044	-0,053	0,021	0,029	0,024	-0,022	-0,011	-0,029	0,006	0,023	0,123	-0,160	0,025	-0,025	0,173	-0,013	-0,472	1,000	-0,182	-0,249	-0,198	-0,185
Occtype3	0,147	0,062	0,077	0,041	-0,027	-0,014	-0,028	0,055	0,065	0,087	-0,043	-0,072	-0,082	0,057	-0,078	-0,068	0,011	-0,130	0,105	-0,018	-0,016	-0,098	0,064	-0,357	-0,182	1,000	-0,188	0,138	0,141
Occtype4	0,165	0,181	0,093	0,003	-0,020	-0,019	0,024	0,029	0,051	-0,002	-0,049	0,007	-0,008	-0,034	0,022	0,015	0,017	-0,163	0,115	-0,001	0,008	-0,121	0,080	-0,488	-0,249	-0,188	1,000	0,164	0,247
TENUEXPE	0,936	0,544	0,157	0,035	-0,060	-0,044	0,028	-0,280	0,037	0,016	-0,059	-0,020	0,036	0,077	0,074	-0,016	-0,107	-0,013	0,328	0,075	0,229	-0,453	-0,132	-0,066	-0,198	0,138	0,164	1,000	0,285
Ln (WAGEHRPP)	0,326	0,215	0,143	-0,012	0,000	-0,002	0,019	0,216	-0,007	0,063	0,038	-0,041	-0,098	0,037	-0,006	-0,038	-0,017	-0,156	0,211	0,028	-0,007	-0,214	0,085	-0,145	-0,185	0,141	0,247	0,285	1,000

Bibliographie

BAYET A. (1996), *L'éventail des salaires et ses déterminants*, Données sociales.

BEVERS T., COLLARD A., DE SPIEGELEIRE M., DE VOS D., GILBERT V. et VAN HOVE H. (2010), *L'écart salarial entre les femmes et les hommes en Belgique - Rapport 2010*, Institut pour l'égalité des femmes et des hommes.

BROWN R., MOON M. et ZOLOTH B. (1980), *Incorporating occupational attainment in studies of male-female earnings differentials*, The Journal of Human Resources, 15 (1), pp.3-28.

DE CURRAIZE Y. et HUGOUNENG R. (2004), *Inégalités de salaires entre femmes et hommes et discrimination*, Revue de l'OFCE (Observatoire Français des Conjonctures Economiques), pp. 193-224.

GLAUDE M. (1987), *Les salaires dans la fonction publique*, Données sociales.

L'HERITIER J.-L. (1992), *Les déterminants du salaire*, Economie et Statistique, n° 257

MEURS D., PONTHIEUX S. (2000), *Une mesure de la discrimination dans l'écart de salaire entre hommes et femmes*, Economie et Statistique, n° 337-338.

MINCER J. et POLACHEK S. (1974), *Family Investment in Human Capital: Earnings of Women*, Journal of Political Economy, vol. 82, n° 2

OAXACA R. L. (1973), *Male-Female Wage Differentials in Urban Labour Markets*, International Economic Review, 14, pp. 693-709.

SOFER C. (1990), *La répartition des emplois par sexe: capital humain ou discrimination?*, Economie et prévision, n° 92-93.

THIRY B. (1985), *La discrimination salariale entre hommes et femmes sur le marché du travail en France*, Annales de l'INSEE¹³, n° 58.

TIJDENS K.G. et OSSE P., *WageIndicator continuous web-survey on work and wages*, University of Amsterdam, Amsterdam Institute for Advanced Labour Studies (AIAS) and StichtingLoonwijzer

TIJDENS K.G., VAN ZIJL S., HUGHIE-WILLIAMS M., VAN KLAVEREN M. et STEINMETZ S. (2010), *Codebook and explanatory note on the WageIndicator dataset, a worldwide, longitudinal, multilingual web and paper survey on work and wages*, University of Amsterdam, Amsterdam Institute for Advanced Labour Studies (AIAS).

VAN LAERE Ch. (2000), *Evaluation et classification de fonctions. Des outils pour l'égalité salariale*, Ministère fédéral de l'Emploi et du Travail, Direction de l'égalité des chances (Belgique).

¹³ INSEE signifie Institut National de la Statistique et des Etudes Economiques