

**Y A-T-IL VRAIMENT DISCRIMINATION SALARIALE
CONTRE LES FEMMES ?**

Laure Dutoit
avec l'appui de Jean-Christian Lambelet

R é s u m é

Etablir scientifiquement l'existence et l'importance d'une éventuelle discrimination salariale selon le sexe se heurte à des difficultés conceptuelles et techniques dont la portée n'est pas toujours reconnue : notre analyse d'un grand échantillon de données pour la Suisse en 1997 aboutit à la conclusion que – *selon l'approche utilisée* – il y a discrimination significative contre les femmes ; ou que l'on ne peut détecter de discrimination ; ou qu'il y a discrimination significative contre...les hommes. Les approches utilisées sont : la régression directe générale (pour des qualifications et autres attributs donnés, les salaires sont-ils différents selon le sexe ?) ; la régression inverse générale (pour un salaire donné, les employeurs demandent-ils des qualifications et autres attributs différents selon le sexe ?) ; la méthode d'Oaxaca-Blinder (l'effet d'un attribut – l'âge ou l'état civil, par exemple – sur le salaire peut différer selon le sexe et il faut donc estimer et comparer des équations séparées pour les femmes et pour les hommes) ; la correction de Heckman (le fait de se concentrer sur les seules personnes employées peut engendrer un biais de sélection). Notre conclusion générale est agnostique : dans l'état actuel des techniques et des connaissances, il n'est pas (encore ?) possible de se prononcer scientifiquement sur l'existence ou non, en Suisse, d'une discrimination salariale fondée sur le sexe. Cette conclusion se démarque donc de celle des études préexistantes, lesquelles tendaient à établir l'existence d'une discrimination contre les femmes.

A b s t r a c t

Trying to ascertain scientifically the existence and importance of gender-based salary discrimination runs into conceptual and technical difficulties, the implications of which are not always fully appreciated : *depending on the approach used*, our analysis of a large sample for Switzerland in 1997 concludes that there is significant discrimination against women ; or that no discrimination can be detected ; or that there is significant discrimination against...men. The approaches used are : direct general regression (for given attributes such as qualification levels, do salaries differ according to gender ?) ; inverse general regression (for a given salary level, do employers require different attributes, such as qualification levels, depending on gender ?) ; the Oaxaca-Blinder method (the effect of a given attribute – e.g. age or marital status – on salaries may differ depending on gender and one must therefore estimate and compare separate equations for women and for men) ; Heckman's correction (focusing on working persons only may give rise to a selection bias). Our general conclusion is agnostic : given the present state of techniques and knowledge, it is not (yet ?) possible to ascertain scientifically the existence of gender-based salary discrimination in Switzerland. This conclusion differs from that reached by pre-existing studies, which tended to underpin the existence of salary discrimination against women.

Keywords : gender-based salary discrimination ; direct and inverse regressions ; the Oaxaca-Blinder approach ; Heckman's correction ; Switzerland.

JEL Classification : J71 ; J31 ; D63 ; C29

Y a-t-il vraiment discrimination salariale contre les femmes ?

Laure Dutoit¹

avec l'appui de Jean-Christian Lambelet

1. Introduction

1.1 Problèmes conceptuels et techniques : un premier aperçu

Les études sur le marché du travail en Suisse et dans beaucoup d'autres pays relèvent sans exception que les femmes ont *en moyenne* des niveaux de salaire inférieurs à ceux des hommes. A lui seul, ce fait d'observation ne suffit cependant pas à établir l'existence d'une discrimination salariale contre les femmes parce que les différences de salaire constatées pourraient tenir à des différences de *productivité*, elles-mêmes dues à divers facteurs tels que des niveaux de formation différents. En d'autres termes, il y a discrimination si, *pour un niveau de productivité donné*, le salaire des femmes est inférieur à celui des hommes.²

Éliminer l'effet des différences de productivité pour vérifier s'il reste ou non des écarts salariaux dus à une discrimination se heurte cependant à des difficultés conceptuelles et techniques qui ne sont pas toujours connues ou reconnues. Deux exemples simples peuvent l'illustrer à titre d'introduction.

Un premier exemple est dû à Conway-Roberts (1983). Soient les données schématiques et fictives suivantes, lesquelles sont fortement simplifiées à des fins d'illustration :

Tableau 1
Nombre de personnes travaillant dans l'entreprise X

	Hommes			Femmes		
	Type d'emploi					
	Employé	Cadre	Total	Employé	Cadre	Total
Formation						
Maturité	40	0	40	80	0	80
Licence	23	17	40	16	4	20
Master	1	19	20	0	0	0
Total	64	36	100	96	4	100

On constate que 80 femmes sur 100 travaillant dans l'entreprise X ont arrêté leur formation au niveau de la maturité, contre 40 hommes sur 100. 20 femmes sur 100 sont allées jusqu'à la licence, contre 40 hommes sur 100. Enfin, 20 hommes sur 100 ont un master alors que ce n'est le cas d'aucune des femmes. Les hommes ont donc un niveau moyen de formation plus élevé que les femmes. Par ailleurs, on admet que la fonction de cadre est préférable à celle d'employé, ne serait-ce qu'en raison d'une rémunération plus élevée.

¹ DEEP/HEC/UNIL. Cette étude repose sur un travail de séminaire effectué et présenté pendant l'année universitaire 2000/2001 dans le cadre du cours de modèles macroéconomiques du Prof. J.-Ch. Lambelet. Ce dernier en a remanié la forme pour les besoins de la publication, en y ajoutant quelques inputs de son cru. La version originale du travail de séminaire est disponible sur : <http://www.hec.unil.ch/modmacro/>. Pour l'aide qu'ils nous ont généreusement accordée, nos remerciements vont particulièrement à Enrico Bolzani, Jacques Huguenin, Christophe Kolodziejczyk, Alexander Mihailov et Délia Nilles, avec une mention toute spéciale pour José Anson, assistant du cours.

² "Productivité" est toujours pris ici dans le sens de "productivité marginale".

Une première manière d'analyser cet échantillon pour déterminer s'il y a discrimination selon le sexe est de calculer la distribution des types d'emploi selon le niveau de formation.

Tableau 2
Pourcentages dans le sens horizontal

	Hommes			Femmes		
	Type d'emploi					
	Employé	Cadre	Total	Employé	Cadre	Total
Formation						
Maturité	100%	0%	100%	100%	0%	100%
Licence	57,5%	42,5%	100%	80%	20%	100%
Master	5%	95%	100%	0	0	-
Total	64%	36%	100%	96%	4%	100%

Examinons les lignes horizontales de ce tableau, soit la distribution par niveau de formation.³ On constate ainsi qu'il n'y a aucune discrimination pour celles et ceux qui se sont arrêtés au niveau de la maturité : 100% des hommes et des femmes qui ont uniquement une maturité sont des employés. En revanche, 42,5% des hommes qui sont allés jusqu'à la licence ont un emploi de cadre alors que ce n'est le cas que de 20% des femmes. On conclut donc qu'il y a "clairement" discrimination à l'encontre des *femmes*.⁴

Une deuxième manière d'analyser l'échantillon pour déterminer s'il y a discrimination selon le sexe est de calculer la distribution des niveaux de formation selon le type d'emploi (alors que c'était l'inverse dans le tableau précédent, lequel reproduit la distribution des types d'emploi par niveau de formation).

Tableau 3
Pourcentages dans le sens vertical

	Hommes			Femmes		
	Type d'emploi					
	Employé	Cadre	Total	Employé	Cadre	Total
Formation						
Maturité	62,5%	0%	40%	83,3%	0%	80%
Licence	35,9%	47,2%	40%	16,7%	100%	20%
Master	1,6%	52,8%	20%	0%	0%	0%
Total	100%	100%	100%	100%	100%	100%

Le tableau se lit maintenant dans le sens vertical.⁵ Comme précédemment, on constate que, parmi celles et ceux qui occupent un poste de cadre, personne – homme ou femme – ne s'est arrêté au niveau de la maturité. A ce niveau-là, il n'y a donc pas de discrimination. En revanche, 16,7% des employées (féminin) ont une licence et 83,3% n'ont qu'une maturité alors que ces proportions se montent à 35,9% et 62,5% chez les employés (masculin). Symétriquement, 47,2% seulement des hommes qui sont allés jusqu'à la licence ont une fonction de cadre, contre 100% pour les femmes. On conclut donc maintenant qu'il y a "clairement" discrimination à l'encontre des *hommes*.

³ On note que les sommes totales (100%) sont faites dans le sens horizontal.

⁴ On ne peut rien dire au niveau du master, puisque l'échantillon ne comprend aucune femme en ayant un.

⁵ Les sommes totales (100%) sont maintenant faites dans le sens vertical.

Cette différence d'optique ou d'interprétation se retrouvera plus loin lorsque nous comparerons les résultats des régressions directes et ceux des régressions indirectes.

Un deuxième exemple, dû à Weisberg-Tomberlin (1982), permet de comprendre ce qui est en cause. Cet exemple illustre aussi une des difficultés ou un des pièges qui existent lorsqu'on cherche à déterminer après coup si une entreprise a suivi une politique d'embauche discriminatoire ou non. Nous allons le paraphraser de la manière suivante.

Supposons qu'une entreprise donnée souhaite sélectionner six personnes pour un certain type d'emploi E . Dix femmes et dix hommes se présentent. L'entreprise leur fait passer une batterie de tests dont on va admettre, pour les seuls besoins de l'argumentation, qu'ils permettent de mesurer parfaitement, sur une échelle de 0 à 100, la productivité future P des candidates et candidats – il n'y a donc pas d'incertitude quant à leur "valeur" pour l'entreprise. Voici les résultats des tests selon le sexe des candidats et dans un ordre ascendant :

Hommes : $P = 40, 50, 60, 60, 70, 75, 80, 90, 95, 95$ (moyenne : 71,5)

Femmes : $P = 30, 35, 55, 60, 60, 75, 75, 75, 85, 90$ (moyenne : 64,0)

En moyenne, les candidats sont donc plus productifs que les candidates, ce qui pourrait tenir, comme précédemment, à des niveaux de formation différents et, plus généralement, à des qualifications différentes.

En appliquant le critère parfaitement *non discriminatoire* qui consiste à engager les six personnes les plus productives, et cela indépendamment de leur sexe, l'entreprise embauchera donc le groupe suivant composé de quatre hommes et de deux femmes :

Groupe engagé : $P = 80(\text{h}), 85(\text{f}), 90(\text{h}), 90(\text{f}), 95(\text{h}), 95(\text{h})$

Bien que le critère de sélection appliqué soit strictement le même pour les hommes et pour les femmes, la distribution de la productivité P par sexe n'est pas la même. En particulier, la productivité moyenne est de 90,0 pour les candidats engagés et de 87,5 pour les candidates retenues. Un chercheur qui examinerait la politique d'embauche de l'entreprise après que les engagements ont eu lieu et sur la seule base des données pour P , pourrait donc conclure qu'il y a eu discrimination contre les hommes alors que, par hypothèse, ce n'est pas le cas.⁶ Il serait facile de construire un autre exemple dans la même veine, mais qui déboucherait sur une apparente discrimination contre les femmes.

Plus généralement, ces deux exemples montrent que la probabilité conditionnelle de E (le fait d'obtenir un emploi d'un type donné dans le deuxième exemple ou le type d'emploi dans le premier) étant donné P (la productivité ou, en particulier, les qualifications) ne dépend pas que de la politique d'embauche de l'entreprise, mais qu'elle est aussi déterminée par la distribution préexistante des qualifications dans chaque groupe.

A noter cependant un trait commun aux deux exemples. Dans le premier, il n'y a que deux catégories d'emploi – employé ou cadre – et seulement trois niveaux de formation – maturité, licence, master. Dans le deuxième exemple, la question est de savoir qui sera engagé ou non pour un emploi d'un type donné. Dans les deux exemples, il s'agit donc de variables dites *discrètes* (binaires ou "trinaires"). Supposons maintenant que, dans le deuxième exemple, l'hypothèse d'une éventuelle discrimination soit testée non pas selon que les candidates et candidats ont été engagés ou non, mais selon le salaire de celles et ceux qui ont été embauchés, où

⁶ L'existence d'une discrimination (ici apparente) n'est évidemment pas qu'une question de moyennes et il faudrait encore examiner les autres moments de la distribution.

les salaires peuvent être considérés comme une variable *continue*. Supposons en outre que les salaires soient déterminés uniquement par les productivités individuelles. Il est alors évident, pour ainsi dire par définition, qu'il n'y aura pas de discrimination entre hommes et femmes. Il en ira de même dans le premier exemple si les variables "type d'emploi" et "niveau de formation" peuvent être considérées comme formant un continuum : dans la mesure où les types d'emploi et donc les salaires sont déterminés par les productivités individuelles, ce qui est la question cruciale, il n'y aura alors pas non plus d'indication de discrimination.

Par la suite, nous allons montrer que des difficultés et des pièges du genre de ceux illustrés par ces deux exemples surgissent également lorsqu'on applique les méthodes et outils de *l'économétrie* à un grand échantillon concernant la Suisse en 1997 afin de déterminer s'il y a ou non discrimination salariale selon le sexe. Auparavant, quelques considérations d'ordre général peuvent être utiles.

1.2 Discrimination salariale et économie de marché

Le phénomène de discrimination salariale selon le sexe est-il compatible avec une économie de marché fonctionnant correctement ? La question, assez fondamentale, mérite sans doute qu'on s'y arrête un moment. Comme on va le voir, notre réponse sera nuancée.

Supposons que, dans une société non libérale au plan politique – cela pourrait être l'Afrique du Sud au temps de l'apartheid ou les États du sud des USA jusqu'à assez récemment –, *tous* les employeurs souffrent d'un fort préjugé à l'encontre d'une catégorie de la population (les Noirs, en l'occurrence) qu'ils considèrent comme "inférieurs", y compris au plan économique. Agissant en quelque sorte comme un *cartel*, ils paieront alors aux Noirs des salaires réels inférieurs à leur productivité marginale, la différence étant tout bénéfique pour eux.⁷ Tant qu'il y a un excédent d'offre sur le marché du travail ou, à la limite, tant qu'on est à l'équilibre de plein emploi, il n'y a pas d'incitant pour les membres du cartel à "faire défection" (c'est-à-dire à payer des salaires plus élevés aux Noirs). Mais l'incitant est présent dès qu'il y a un excédent de demande sur le marché du travail, comme c'est généralement le cas en période de haute conjoncture. Avec des phases de haute conjoncture survenant de manière plus ou moins régulière, on pourrait donc arguer que la logique de l'économie de marché devrait mener à une érosion plus ou moins progressive de la discrimination et cela peut-être jusqu'à sa disparition complète. Cela pourrait cependant prendre beaucoup de temps, d'autant plus qu'on peut imaginer que les membres du cartel qui seraient tentés de faire défection pourraient faire l'objet de pressions directes d'ordre physique ou moral, c'est-à-dire non économique. Bref, le raisonnement comme l'expérience historique suggèrent que la discrimination salariale peut exister dans une économie de marché, peut-être pendant fort longtemps, mais que cette économie de marché ne sera pas de type concurrentiel, la discrimination reposant sur un comportement cartellaire.

A noter toutefois qu'une authentique discrimination salariale peut aussi avoir des fondements beaucoup plus honorables que dans les cas ci-dessus. En Suisse, l'État social était jusque dans les années 1950 sinon inexistant, du moins embryonnaire (l'AVS date de 1947-48). A l'époque, il était généralement admis que les hommes avaient typiquement charge de famille, que le destin "normal" des femmes était de se marier, puis de s'occuper à plein temps de leur famille et de leurs enfants, alors que ce n'était généralement pas le cas des femmes qui travaillaient hors de l'économie familiale – *ergo* il convenait, par consensus général et tradition,

⁷ Il y a donc *concertation* – explicite ou tacite – entre les employeurs pour ne pas payer la main-d'œuvre noire plus qu'une certaine fraction de leur productivité, la concertation devant également porter sur la taille de cette fraction. C'est cet élément de concertation qui nous incite à parler de cartel.

de payer les hommes plus que les femmes, même à occupation et qualifications égales, et cela pour des raisons de justice (re)distributive sans rapport avec un éventuel sexisme primitif. En lieu et place d'un État social officiel, on pouvait donc se trouver en présence d'un équivalent au moins partiel, mais non étatique. Cette discrimination salariale était aussi plus honorable que dans les cas cités ci-dessus dans la mesure où elle consistait à payer les hommes plus et les femmes moins que leur productivité marginale, de sorte qu'on se trouvait en présence d'un système de transferts sociaux gérés par les employeurs. Était-ce le cas ? Comme on ne peut mesurer directement les productivités marginales, aujourd'hui et a fortiori naguère, la question doit rester sans réponse assurée et, à notre connaissance, personne ne l'a jamais soulevée ni examinée. Rien ne permet cependant d'exclure une réponse positive. Dans ce cas, la montée de l'État social moderne et l'évolution concomitante des mentalités se seraient accompagnées d'une disparition progressive de son équivalent non étatique – et donc d'une disparition progressive de la discrimination salariale selon le sexe. S'il devait néanmoins subsister une certaine discrimination aujourd'hui, ce qui n'est pas prouvé comme on le verra plus loin, on pourrait alors y voir un vestige évanescent de l'ancien État social et des attitudes éthiques qui le sous-tendaient.

1.3 Discrimination salariale et justice sociale

Il est des emplois qui ne demandent pratiquement pas de qualification ou que très peu. Un exemple est la fonction de caissier ou caissière dans une grande surface moderne, laquelle consiste aujourd'hui essentiellement à faire passer des objets de consommation sur un lecteur optique, la caisse enregistreuse s'occupant de tout le reste, sauf de rendre la monnaie (et encore n'est-ce pas nécessaire en cas de paiement par carte). Si, dans ces circonstances, on devait observer que caissiers et caissières ne sont pas rémunérés de façon strictement égale par heure de travail, on se trouverait assurément en présence d'un cas de discrimination salariale caractérisé.

Très différent est le cas des emplois demandant une ou des qualifications poussées. Prenons l'exemple d'une banque ou institution analogue qui veut engager un jeune analyste financier (homme ou femme). Pour cet employeur, la productivité et donc la rémunération dudit analyste n'est, à la différence du caissier ou de la caissière cité ci-dessus, pas qu'une affaire de performance par heure de travail *hic et nunc*. Même s'il sort d'une haute école prestigieuse, il devra généralement être formé par l'entreprise et cela largement aux frais de cette dernière.⁸ Pour elle, il sera alors davantage que de la simple main-d'œuvre, mais deviendra en fait une partie de son capital. Ce capital est cependant d'un type particulier, en cela qu'à la différence du capital "ordinaire" (des machines, par exemple) il ne peut faire l'objet d'une appropriation par l'entreprise, ou que très partiellement,⁹ de sorte qu'en cas de départ prématuré de l'analyste elle enregistrera une perte sur ce qui était en fait un investissement dans le capital humain dudit analyste. Pour l'entreprise, l'engagement de cette personne est donc, dans une certaine mesure, un pari sur l'avenir et la décision de l'engager reposera en partie sur des considérations de probabilité : l'analyste tiendra-t-il ses promesses s'il est engagé *et* formé par l'entreprise ? et plus spécifiquement, restera-t-il assez longtemps en son sein pour que l'investissement consenti pour le former puisse être amorti ? L'entreprise sera d'autant plus portée à investir dans la formation dudit analyste que ces conditions ont de bonnes chances d'être satisfaites.

⁸ Cette formation peut être interne, c'est-à-dire du type *on the job*, et/ou se faire en dehors de l'entreprise (cours dans une institution spécialisée, etc.)

⁹ "Que très partiellement" parce que l'entreprise peut se prémunir dans une certaine mesure contre un départ prématuré – et donc contre une perte de son capital – au moyen d'un contrat de travail à long terme, avec pénalité en cas de départ prématuré, engagement de ne pas travailler pour la concurrence pendant un certain temps, etc.

Le fait biologique que ce sont les femmes qui mettent les enfants au monde, et qu'il est généralement admis que leur présence plus ou moins continue est importante dans les premières années de ces derniers, signifie que les probabilités en question seront différentes du point de vue de l'entreprise selon que le jeune analyste en question est d'un sexe ou de l'autre. S'il s'agit d'une femme, il y a un "risque" – du point de vue l'entreprise ! – qu'elle se marie, ait des enfants et les élève pendant quelques années, quittant pour cela l'entreprise ainsi, d'ailleurs, que son métier.¹⁰ Si, malgré ce risque, l'entreprise décide néanmoins d'engager et de former cette jeune analyste, un salaire inférieur à un alter ego masculin correspondra à une décision rationnelle, la différence de rémunération représentant une sorte de prime d'assurance contre le risque d'un départ prématuré et donc d'un investissement non entièrement rentabilisé.

Doit-on appeler cela de la discrimination salariale ? La réponse nous semble clairement négative, dans la mesure où la différence de rémunération se fonde sur des considérations objectives de probabilité et non pas sur un quelconque préjugé sexiste.

Il ne s'ensuit cependant pas que cette situation soit forcément satisfaisante d'un point de vue éthique. Les femmes n'ont pas choisi de naître en tant que telles et elles ne sont pas non plus responsables du fait biologique mentionné plus haut. Ladite différence de rémunération peut donc ne pas constituer de la discrimination salariale au sens où nous l'avons définie et néanmoins être incompatible avec une notion de justice sociale fondée sur l'idée d'une égalité des chances de départ. Il s'ensuit que cette différence de rémunération ne sera pas acceptable d'un point de vue éthique et demandera à être atténuée ou, idéalement, supprimée. Cela peut se faire en légiférant, c'est-à-dire en imposant aux entreprises de verser des salaires égaux "à travail égal", ce qui soulève cependant des problèmes non négligeables d'application dans la pratique et de vérification. Ou cela pourrait se faire au moyen d'un système de transferts, avec une caisse de compensation étatique sur le modèle de celle qui, pendant la dernière guerre mondiale, a assuré une péréquation au moins partielle entre hommes mobilisés et non mobilisés, le problème étant alors entre autres celui du coût d'une bureaucratie supplémentaire. Le choix entre ces deux systèmes, et peut-être d'autres encore, est donc avant tout une question de leur efficacité et de leur coût relatifs.

La différence que nous faisons entre "discrimination salariale", d'une part, et "justice sociale", de l'autre, nous paraît fort importante pour une raison au moins : s'il devait s'avérer qu'il n'y a pas vraiment, au sens précité, de discrimination salariale significative contre les femmes ou, à tout le moins, s'il devait se trouver qu'il n'y a pas d'indications probantes à cet égard, il ne s'ensuivrait pas nécessairement que, dans ce domaine, "tout est pour le mieux dans la meilleure des sociétés non sexistes".

¹⁰ A cela on oppose quelquefois le contre-argument qu'en raison de leurs obligations militaires les hommes aussi seront, au cours de leur carrière, absents de leur travail pendant des périodes prolongées. Cependant, ces périodes sont nettement moins longues que celles liées à une ou plusieurs maternités ; en outre, elles peuvent généralement être aménagées dans une certaine mesure selon les besoins de l'entreprise. Ce contre-argument atténue donc la portée de notre raisonnement, mais ne l'annule pas.

2. Y a-t-il discrimination salariale contre les femmes en Suisse ?

2.1 Données utilisées

Pour tenter de répondre à la question en titre de cette section, nous utilisons un échantillon tiré de l'*Enquête suisse sur la population active* (ESPA) effectuée en 1997. Cette enquête a porté sur quelque 16 000 ménages. Le nombre d'observations utilisables est cependant nettement plus petit, car on élimine – du moins dans un premier temps¹¹ – les chômeurs et les personnes ne faisant pas partie de la population active, c'est-à-dire les personnes n'ayant pas reçu de salaire dans la période couverte par l'échantillon. Ceci fait tomber l'échantillon à 9 922 observations. En outre, toutes ces personnes n'ont pas pu ou voulu répondre à toutes les questions pertinentes de notre point de vue, de sorte que les données les concernant ne peuvent être utilisées. En fin de compte, cela laisse 3 666 observations utilisables, ce qui constitue néanmoins un grand échantillon.

Par sa nature, cet échantillon n'est pas différent de celui (fictif) qui figure au tableau 1 ci-dessus, à cela près qu'au moins certaines variables sont continues (les salaires), ou quasiment continues (l'âge mesuré en années), et que l'échantillon est aussi beaucoup plus riche en information. Outre le revenu annuel brut du travail, c'est-à-dire la variable centrale du point de vue de la discrimination salariale, toute une série d'attributs a en effet été relevée, des attributs qui peuvent donc être inclus dans l'analyse : niveau de formation, sexe, âge, état civil, type de contrat de travail, ancienneté dans l'emploi, niveau hiérarchique, nationalité, etc.¹²

2.2 Problématique générale

Notre approche est celle de l'économétrie appliquée à un échantillon en coupe transversale. La plupart des études économétriques qui ont été faites dans ce contexte, en particulier celles pour la Suisse [voir la bibliographie dans Donzé (2000) et, tout particulièrement, Diekmann-Engelhardt (1994) ainsi que Flückiger-Ramirez (2000)], suivent une approche "standard" que nous allons discuter sous une forme simplifiée, en ce sens que nous admettons pour un moment que le salaire ne dépend que de la qualification et du sexe.

Soit l'équation :

$$\log(\text{salaire}_i) = a + b \text{ qualification}_i + c \text{ sexe}_i + e_i \quad (i = 1, \dots, 3\,666) \quad (1)$$

La variable salaire_i est le revenu brut du travail de la personne i .¹³ La variable "qualification $_i$ " est un indice discret mesurant le degré de qualification de la personne i et la variable binaire "sexe $_i$ " indique son sexe (0 = homme ; 1 = femme). Enfin, " e_i " est un terme aléatoire rendant compte des erreurs possibles de mesure affectant la variable dépendante ainsi que de tous les autres facteurs déterminant le salaire, mais dont aucun n'a – ici par hypothèse – d'influence significative sur le salaire.¹⁴

Dans cette équation, c'est le paramètre c qui mesure l'existence et la taille d'une éventuelle discrimination salariale selon le sexe : s'il n'est pas significativement différent de zéro, il n'y a pas de discrimination ou, plus exactement, pas de discrimination discernable dans l'échantil-

¹¹ La correction dite de Heckman conduit en effet à utiliser les observations pour les personnes au chômage et les inactifs – voir plus loin.

¹² La liste complète des attributs se trouve au début de l'*Appendice*.

¹³ Comme dans la plupart des autres études, on prend le logarithme du salaire, en particulier parce que cela facilite l'interprétation des coefficients b et c .

¹⁴ La constante a n'a pas d'importance (il est toujours possible de la rendre nulle en changeant les unités de mesure des variables).

lon ; s'il est significativement différent de zéro et négatif, il y a discrimination salariale contre les femmes : à qualification égale, les hommes touchent un salaire plus élevé ; enfin, si c est significativement différent de zéro et positif, il y a discrimination salariale contre les hommes.

Comme dit plus haut, l'équation (1) est en essence celle utilisée dans la plupart des études économétriques, à cela près que d'autres attributs sont pris en considération en sus de la qualification et du sexe. Les paramètres b et c , ainsi que ceux des autres attributs retenus, sont estimés économétriquement, par exemple au moyen de la méthode de régression dite des moindres carrés ordinaires, et ils sont testés quant à leur signe et degré de signification. Les études existantes pour la Suisse, en particulier celles de Diekmann-Engelhardt (1994) et de Flückiger-Ramirez (2000), arrivent ainsi à la conclusion qu'il y a une discrimination statistiquement significative *contre les femmes*.

Cette approche "standard" se heurte toutefois à un certain nombre de difficultés conceptuelles et statistiques qui ne sont pas toujours connues ou reconnues.

Une *première* difficulté est celle qu'on a vue dans les deux exemples discutés plus haut : on ne tient pas compte de la distribution préexistante de la variable discrète "qualifications", ce qui peut conduire à des résultats spacieux ("*spurious*"), comme ces exemples le montrent.

Une *deuxième* difficulté, qui constitue un problème séparé, tient à la méthode d'estimation économétrique, telle par exemple les moindres carrés ordinaires. Pour voir de quoi il retourne, notons que l'équation (1) et l'équation suivante sont strictement équivalentes :

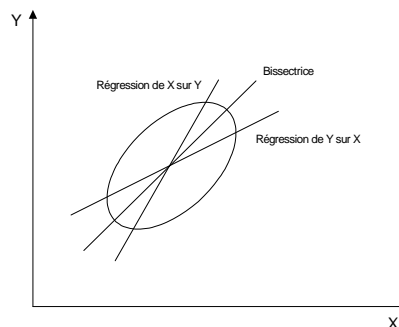
$$\text{qualification}_i = - (a/b) + (1/b) \log(\text{salaire}_i) - (c/b) \text{sexe}_i + e^*_i \quad (2)$$

où $e^*_i = - (1/b) e_i$.

Pour la commodité de l'exposé, nous appellerons l'équation (1) la "régression directe" et l'équation (2) la "régression inverse". L'équation (1) dit que le salaire fixé par l'employeur est déterminé par la qualification et par le sexe : étant donné tel ou tel degré de qualification et étant donné le sexe, l'employeur paiera en probabilité tel ou tel salaire. L'équation (2), quant à elle, dit que pour un niveau de salaire et un sexe donnés, l'employeur demandera en probabilité telle ou telle qualification. En logique, les deux interprétations sont strictement équivalentes, comme sont mathématiquement équivalentes les équations (1) et (2).

Si elles sont mathématiquement équivalentes, l'équation (1), soit la régression directe, et l'équation (2), soit la régression inverse, ne le sont cependant pas *en économétrie*, comme le graphique 1 le montre. Étant donné que le graphique est en deux dimensions, il concerne le cas d'une équation générale à deux variables seulement, Y et X, mais il permet d'illustrer commodément le problème sous discussion.

Graphique 1



On voit ainsi que – comme cela est bien connu – la régression directe $Y = \alpha + \beta X + \varepsilon$ donnera des résultats (paramètres estimés) différents de la régression inverse $X = \alpha' + \beta' Y + \varepsilon'$. En particulier, les valeurs estimées des pentes β et β' ne satisferont pas la condition $\beta = 1/\beta'$. Comme on sait, cela tient au fait que la minimisation des écarts au carré se fait différemment – verticalement dans un cas, horizontalement dans l'autre. Le fond du problème est que les équations (1) et (2) ne sont pas des *fonctions*, mais des *relations* ; ou, si l'on préfère, il n'est pas clair si la direction de la causalité va de la qualification au salaire ou si elle va dans l'autre sens. Or, les méthodes économétriques tels les moindres carrés sont destinées à estimer des fonctions, non des relations.¹⁵

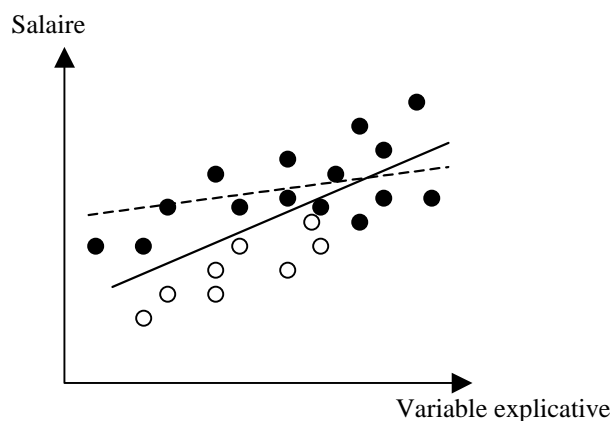
A noter que ce problème se pose même si la première difficulté mentionnée plus haut n'existe pas : en admettant que la distribution préexistante des qualifications soit la même pour les hommes et les femmes et que cette variable soit continue, la question de la direction de la régression subsisterait, avec des résultats différents selon que l'on choisit la régression directe ou inverse.

Cela peut avoir des conséquences cruciales au plan de l'interprétation de ces résultats. Dans le graphique ci-dessus, supposons que la pente de la bissectrice corresponde à l'absence de discrimination salariale, qu'une pente plus raide indique une discrimination contre les hommes et une pente moins raide une discrimination contre les femmes. La régression directe révélerait alors une discrimination contre les femmes, mais la régression inverse indiquerait une discrimination contre les hommes. C'est exactement ce que nous allons constater avec l'échantillon suisse, comme on verra plus loin.

Une *troisième* difficulté a été mise en lumière par Oaxaca-Blinder, comme explicité dans Flückiger-Ramirez (2000). Prenons le cas d'une équation qui inclut l'âge de la personne parmi les attributs utilisés comme variables explicatives des salaires. Pour les raisons énoncées plus haut, une entreprise rationnelle et "non sexiste" considérera différemment les perspectives futures d'un jeune analyste financier selon qu'il s'agira d'un homme ou d'une femme et, par conséquent, pourra leur offrir des salaires différents. Par conséquent, le paramètre de la variable "âge" peut avoir une valeur différente selon qu'il s'agit d'une femme ou d'un homme. La méthode d'Oaxaca-Blinder consiste dès lors à estimer non pas une équation générale pour les hommes et les femmes, mais deux équations, l'une pour les hommes et l'autre pour les femmes, puis à tester les différences entre les paramètres de chacune des deux équations.

Une *quatrième* et dernière difficulté tient à la possibilité d'un biais de sélection dans l'échantillon. Elle peut être représentée ainsi :

Graphique 2



¹⁵ Pour plus de détails, voir Lambelet-Mihailov (2001).

Lorsqu'on régresse le salaire sur les variables explicatives, on n'inclut généralement pas les inactifs et les chômeurs dans l'échantillon. Il peut – mais il ne doit pas nécessairement – en résulter un biais de sélection dans l'échantillon, comme le graphique 2 le montre. Ce biais de sélection est dû au fait que, par définition, les inactifs et les chômeurs n'ont pas d'emploi rémunéré, peut-être parce que le salaire qu'ils demandent – leur "salaire de réservation" – est supérieur au salaire offert par les entreprises étant donné leurs qualifications (auto-sélection) ; par conséquent, on ne peut observer leurs salaires. L'échantillon utilisé peut ainsi ne pas être représentatif de toute la population. Dans le graphique, les points blancs sont les salaires non observables et les points noirs sont ceux qu'on observe. Si tous les points étaient connus, et dans la mesure où la distribution des points blancs et noirs est à l'image de celle dans le graphique, la "vraie" droite de régression est celle représentée par la ligne pleine. Mais si l'on néglige les inactifs et chômeurs, la droite de régression est celle indiquée en pointillés. Clairement, il y aura alors un biais d'échantillonnage qui conduira à des conclusions incorrectes quant à la présence ou non de discrimination salariale. Pour y obvier, Heckman propose la méthode suivante¹⁶ : premièrement, il faut estimer la probabilité qu'une personne interrogée participe au marché du travail, qu'elle soit active, au chômage ou inactive. Deuxièmement, il faut introduire cette probabilité en tant que variable explicative dans la régression. En procédant de cette façon, on utilise toute l'information à disposition.

Dans les sections suivantes, nous allons présenter, comparer et commenter les estimations obtenues en appliquant à notre échantillon

- (1) la méthode de la régression directe générale,
- (2) celle de la régression inverse générale,
- (3) celle d'Oaxaca-Blinder, et enfin
- (4) la correction de Heckman.

Ce faisant, nous essaierons d'être relativement *peu techniques*. C'est-à-dire que nous nous limiterons à des considérations dont nous espérons qu'elles seront accessibles à des lectrices et lecteurs non rompus à l'économétrie. La liste complète des attributs pris en considération, les résultats détaillés des calculs et régressions ainsi que quelques commentaires plus techniques sont ainsi relégués dans un **appendice** destiné aux spécialistes.

2.3 Régression directe générale (pour plus de détails, voir la section A.2 de l'appendice)

La régression directe "générale" (c'est-à-dire pour l'ensemble des hommes et des femmes) est analogue à l'équation (1) ci-dessus.¹⁷ Avec treize attributs, on arrive à expliquer 53% de la variation dans les salaires annuels bruts (R^2 ajusté = 0,53), ce qui est honorable pour une coupe transversale. Le taux d'occupation¹⁸ est l'attribut le plus significatif ($t = 24,3$), ce qui ne surprend pas, suivi de l'éducation ou formation¹⁹ ($t = 14,5$), ce qui ne surprend pas non plus. Les seuls attributs non significatifs au niveau de 1% sont l'état civil²⁰ (1,5), l'ancienneté de l'employé ou employée dans l'entreprise²¹ (1,6) et son niveau hiérarchique²² (1,1) ainsi que le

¹⁶ C'est une des raisons pour lesquelles il a récemment reçu le Prix Nobel d'économie.

¹⁷ Elle a été estimée par les moindres carrés ordinaires, à cela près que comme il y avait des indications d'hétéroscédasticité, la correction de White a été appliquée, comme aussi pour la plupart des autres équations.

¹⁸ Défini comme le rapport entre le nombre d'heures travaillées par semaine et le nombre moyen (40 heures).

¹⁹ Cette variable discrète est codée ainsi : 1 = école obligatoire ou formation élémentaire; 2 = apprentissage, école professionnelle, école de formation générale ou maturité; 3 = maîtrise ou école technique ; 4 = haute école professionnelle, haute école technique ou université.

²⁰ Variable binaire (1 = marié; 0 = autre).

²¹ Variable binaire (1 = 8 années dans l'entreprise ou plus; 0 = moins de 8 années).

²² Cette variable discrète va de 1 = cadre supérieur à 9 = ouvrier ou employé non qualifié.

secteur (privé ou public) dans lequel l'entreprise employante – firme privée, administration ou autre – est active (0,9).²³ Les attributs suivants ont un effet à la fois positif et significatif sur le salaire annuel brut : la taille de l'entreprise (7,1),²⁴ l'expérience de l'employé ou employée (5,2),²⁵ sa nationalité²⁶ (4,4), la branche économique à laquelle l'entreprise appartient²⁷ (4,4), le type de contrat de travail²⁸ (3,6) et l'âge de l'employé ou employée²⁹ (2,6).

Le coefficient de la variable "sexe de l'employé" (= 0 pour les hommes et = 1 pour les femmes) est négatif et hautement significatif ($t = -5,6$). Cela signifie donc que, selon cette régression directe et générale, il y aurait une discrimination salariale marquée *contre les femmes* : toutes choses par ailleurs égales, les femmes auraient un salaire inférieur d'environ 16% par rapport à celui des hommes.

2.4 Régression directes séparées (pour plus de détails, voir A.3 dans l'appendice)

Dans ce cas, on estime des régressions séparées pour les femmes et les hommes, et cela pour les raisons mentionnées plus haut. Les éventuelles différences entre les coefficients sont ensuite décomposées selon la méthode d'Oaxaca-Blinder. La dissemblance des résultats des deux régressions confirme que l'effet des variables explicatives peut être différent selon le sexe et justifie donc l'application de ladite méthode.

On trouve ainsi que 29,1% de la différence totale entre le salaire des hommes et celui des femmes ne sont pas expliqués par des différences dans les attributs, mais peuvent être attribués à une discrimination salariale *contre les femmes*. Certaines variables – celle de l'éducation ou formation, par exemple – suggèrent cependant une discrimination ponctuelle contre les hommes. Il reste que, dans cette méthode, c'est la différence totale qui indique la direction générale de la discrimination salariale.³⁰

2.5 Régressions inverses générales et séparées avec index des qualifications (voir A.4)

Alors que la régression directe revient à examiner si les hommes et les femmes touchent le même salaire étant donné (par exemple) leurs qualifications, la régression inverse pose la

²³ Ceci va donc à l'encontre d'un a priori qui voudrait qu'il y ait une pratique salariale différente dans les secteurs public et privé.

²⁴ Plus une entreprise est grande et plus elle paie des salaires élevés, toutes choses par ailleurs égales : on peut imaginer que cela est dû à des économies d'échelle et/ou à un plus grand pouvoir de marché pour les grandes entreprises.

²⁵ Cette variable est mesurée par le nombre d'années passées dans le marché du travail sans longue interruption.

²⁶ = 1 pour les Suisse, = 0 sinon. Le fait que le coefficient de cette variable soit significatif et positif signifie-t-il qu'il y a discrimination contre les étrangers ? Cela se pourrait, mais la nationalité pourrait aussi être interprétée comme un signal de stabilité (les Suisses tendent à rester plus longtemps dans l'entreprise).

²⁷ Cette variable vaut 1 si l'entreprise est dans le tertiaire, sinon elle vaut 0. Son paramètre significatif et positif peut être dû au fait que les entreprises du secteur tertiaire tendaient (au moment de l'enquête) à être plus rentables que celles des secteurs primaires ou secondaires.

²⁸ Contrat limité dans le temps = 1, non limité = 2. Un contrat non limité dans le temps peut inciter l'employé ou l'employée à "s'investir" davantage dans l'entreprise.

²⁹ L'âge est mesuré en années rondes. Le fait que cette variable ne soit que faiblement significative pourrait être dû au fait qu'elle entre de façon linéaire dans l'équation. Contre cela, on pourrait imaginer que l'influence de l'âge suive plutôt une courbe en forme de cloche : dans une première phase, l'expérience qui s'accumule l'emporterait sur la perte d'énergie, d'initiative et de créativité qui accompagne en général le passage des années ; dans une deuxième phase, ce serait l'inverse. La question mériterait donc d'être examinée de plus près, par exemple en introduisant l'âge au carré dans l'équation.

³⁰ Les résultats détaillés (voir A.3) révèlent un certain nombre d'autres résultats ponctuels qui sont intéressants. Ainsi, l'état civil reste non significatif pour les femmes, mais devient hautement significatif pour les hommes : il semblerait donc que les employeurs tendent à juger les hommes plus favorablement lorsqu'ils sont mariés. A l'inverse, la nationalité reste significative pour les hommes, mais ne l'est plus pour les femmes.

question dans l'autre sens : pour un salaire donné, les employeurs demandent-ils des qualifications différentes selon le sexe ?

La régression inverse peut se faire de deux manières. Dans cette section-ci, on construit un index des qualifications en utilisant pour cela les coefficients estimés dans la régression directe.³¹ Une deuxième manière, utilisée dans la section suivante, consiste à choisir, comme variable dépendante et à tour de rôle, une des variables composant l'index des qualifications.

Avec la première manière, les résultats pour la régression inverse générale (c'est-à-dire, pour l'ensemble des hommes et des femmes) révèlent maintenant une discrimination hautement significative *contre les hommes* ($t = -18,1$). Si l'on procède selon la méthode d'Oaxaca-Blinder (régression inverse séparée pour chacun des sexes), on trouve que 36,9% de la différence moyenne des qualifications ne sont pas expliqués par les différences entre les moyennes des attributs pris comme variables explicatives, dont le salaire, mais peuvent être attribués à une discrimination contre la gent masculine.

La régression inverse conduit donc à des résultats opposés à ceux de la régression directe. Comme argué plus haut, cela peut tenir à deux causes : premièrement, les variables mesurant les qualifications sont discrètes, ce qui soulève la question des distributions préexistantes de ces qualifications.³² Deuxièmement, cela pourrait aussi être dû à une raison "mécanique", à savoir la direction de la minimisation des erreurs, laquelle est différente selon qu'on applique la régression directe ou inverse. Dans l'état actuel de notre recherche, nous ne pouvons dire laquelle de ces deux causes explique ces résultats contradictoires (les deux pourraient intervenir).³³

2.6 Régressions inverses générales et séparées avec diverses variables dépendantes (voir A.5)

Si l'on prend la variable "éducation-formation" comme variable dépendante, on trouve également une forte discrimination *contre les hommes*, dans le cas de la régression inverse générale comme dans celui des régressions inverses séparées.³⁴ Il en va de même si l'on sélectionne les variables "taux d'occupation" ou "ancienneté". En principe, il est possible de choisir sept variables comme nouvelle variable dépendante. Les résultats pour les quatre autres variables sont cependant trop peu robustes pour qu'on puisse en tirer des conclusions valables.

2.7 La correction de Heckman (pour plus de détails, voir la section A.6 dans l'annexe)

Comme on l'a dit plus haut, la méthode de Heckman permet de détecter la présence d'un éventuel biais de sélection et de corriger ce biais s'il est présent. Comme cette méthode est

³¹ L'ensemble "qualifications" comprend les variables suivantes : âge, éducation-formation, ancienneté, expérience et niveau hiérarchique, auxquelles on ajoute les variables taux d'occupation et type de contrat.

³² A noter que la variable "éducation" soulève un autre problème. L'éducation ou formation est en effet définie selon quatre niveaux codés 1, 2, 3 et 4 (voir la note 19), mais rien ne garantit que le niveau "haute école, université" vaille quatre fois le niveau "école obligatoire".

³³ Étant donné ces résultats contradictoires selon la direction de la régression, on pourrait se demander s'il ne conviendrait pas d'avoir recours à une autre méthode d'estimation que celle des moindres carrés, comme par exemple celle de la régression dite orthogonale. L'ennui, cependant, est que les moments d'ordre supérieur de la distribution des coefficients sont infinis dans cette méthode (Lambelet-Mihailov 2001). Par conséquent, il n'est pas possible de tester les coefficients estimés quant à leur degré de signification.

³⁴ Les R^2 ajustés pour ces régressions sont cependant peu élevés (0,16, par exemple, dans le cas de la régression inverse générale). Ces résultats doivent donc être considérés avec prudence.

assez complexe, la section correspondante de l'annexe inclut des commentaires relativement détaillés. Ici, nous nous limiterons aux résultats généraux qu'on obtient ainsi.³⁵

Il s'avère que la correction de Heckman ne se justifie pas dans le cas des hommes, ce qui s'explique par le fait que leur probabilité de participer au marché du travail est proche de l'unité. Pour les hommes, les résultats avec la correction de Heckman sont donc fort proches de ceux qu'on obtient sans elle. Pour les femmes, en revanche, la probabilité de participer au marché du travail est significativement plus basse et les résultats sont significativement différents selon qu'on utilise la correction de Heckman ou non – il convient donc d'en tenir compte.

Au total, la régression directe générale avec la dite correction débouche sur une discrimination salariale *à l'encontre des femmes*, comme dans le cas de la régression directe générale sans cette correction. La méthode d'Oaxaca-Blinder appliquée aux régressions directes séparées, avec application de ladite correction, montre en revanche qu'il y a discrimination significative *contre les hommes*.

3. Conclusion générale

Étant donné tous ces résultats, notre conclusion générale ne peut être qu'*agnostique* : étant donné l'état actuel des connaissances et techniques, il n'est pas possible d'établir s'il y a ou non discrimination salariale selon le sexe en Suisse – ou du moins cela n'est pas possible sur la base des données détaillées contenues dans ce grand échantillon qu'est l'ESPA. Il reste donc à espérer que de nouvelles techniques, économétriques ou autres, soient rapidement mises au point qui permettent d'obvier aux difficultés et contradictions auxquelles nous nous sommes heurtés.

Cela étant, il convient peut-être de répéter, en le soulignant, quelque chose qui a été dit plus haut. Supposons donc que de nouvelles techniques permettent de (dé)montrer avant trop longtemps qu'il n'y a pas – ou qu'il n'y a plus dans l'économie suisse moderne – de discrimination salariale significative au sens où nous l'avons définie ; supposons en d'autres termes qu'on trouve que les rémunérations sont déterminées uniquement et objectivement par la productivité présente *et* future des hommes et des femmes. Le fait biologique que ce sont les femmes qui mettent les enfants au monde, assurant par là la perpétuation de l'espèce, et le fait qu'il est généralement considéré souhaitable qu'elles puissent s'occuper de leurs enfants en bas âge – ces faits tendent, toutes autres choses égales, à diminuer leur productivité strictement économique si l'on considère cette dernière sur toute l'étendue de leur vie. La justice sociale ne demanderait-elle alors pas que ce "déficit naturel" de productivité soit compensé d'une manière ou d'une autre ? En outre, la fécondité est aujourd'hui tombée au-dessous du taux de remplacement dans la plupart des pays européens, dont la Suisse, ce qui signifie que la "perpétuation de l'espèce" n'est plus assurée, du moins à l'échelle des communautés nationales. Peut-être cela n'est-il pas étranger au fait que ledit déficit naturel de productivité n'est en général pas compensé ou que partiellement.

³⁵ A noter que la correction de Heckman ne peut s'appliquer qu'aux régressions directes – générales ou séparées –, mais non aux régressions inverses.

Références

- Bourbonnais, R. (2000), *Économétrie*, 3^{ème} édition, éditions Dunod, Paris.
- Conway, D. et H. Roberts (1983), "Reverse Regression, Fairness, and Employment Discrimination", *Journal of Business and Economic Statistics*, vol. 1, n°1, pp. 75-85.
- Diekmann, A. et H. Engelhardt (1994), "Einkommensungleichheit zwischen Frauen und Männern. Eine ökonometrische Analyse der Schweizer Arbeitskräfteerhebung", *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, 131 (1), pp. 57-83.
- Donzé, L. (2000), "Application d'une nouvelle mesure de la discrimination salariale à la situation suisse", *Revue suisse d'économie politique et de statistique*, 2000/3, pp. 341-8.
- Flückiger, Y. et J. Ramirez (2000), *Analyse comparative des salaires entre les hommes et les femmes sur la base de la LSE 1994 et 1996*, Observatoire universitaire de l'emploi, Genève.
- Goldberger, A.S. (1984), "Reverse Regression and Salary Discrimination", *The Journal of Human Resources*, XIX, 3, pp.291–318.
- Heckman, J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, 47 (1), pp. 153-161.
- Hungerbühler, M. (1999), "La discrimination salariale des femmes en Suisse", Université de Lausanne.
- Lambelet, J.-Ch. et A. Mihailov (2001), "The Triple-Parity Law", *Analyses & Prévisions*, Institut Créa, HEC/UNIL, édition du printemps 2001.
- Maddala, G.S. (1992), *Introduction to Econometrics*, 2^{ème} édition.
- Revue *Problèmes économiques* (2000), nov., n° 2.687, pp. 1-2.
- Office fédéral de la statistique (1998), *Enquête suisse sur la population active 1997*, Berne.
- Weisberg, H. et T. Tomberlin (1982), "Statistical Evidence in Employment Discrimination Cases", *Sociological Methods and Research*.

APPENDICE

A.1. Liste des variables

Codage (symboles utilisés)	Signification
Variables concernant l'individu	
age	Age
marie	Etat civil
nat	Nationalité
sexe10	Sexe
edu	Formation acquise
Variables concernant l'individu et l'entreprise employante	
ancien	Ancienneté de l'employé(e) dans l'entreprise
contrat	Type de contrat de travail
exp	Expérience de l'employé(e)
hierar	Niveau hiérarchique de l'employé(e)
tauxoccup	Taux d'occupation de l'employé(e)
Variables concernant l'entreprise employante	
div	Branche économique de l'entreprise
public	Secteur de l'entreprise (privé / public)
tailleta	Taille de l'entreprise

Justification de l'utilisation de ces variables explicatives :

1. Variables concernant l'individu :

L'âge influence le salaire par trois biais. Premièrement, il peut être considéré comme une mesure de l'expérience (en général, plus on est âgé, plus on a travaillé et plus on a ou devrait avoir d'expérience). Deuxièmement, l'âge peut être, dans une certaine mesure, considéré comme une mesure de la stabilité de la personne. Souvent, les jeunes employés quitteront plus facilement leur place de travail (si, par exemple, une autre offre plus d'opportunités ou si l'employée veut fonder une famille) alors que les employés plus âgés tendront à préférer la sécurité de leur place de travail et seront peut-être plus réticents à s'investir dans un nouveau emploi. Troisièmement, il y a le fait biologique que l'énergie, l'initiative et la créativité diminuent en général avec l'âge, même si cela peut être partiellement compensé par une plus grande expérience, une plus grande "sagesse" et une meilleure "auto-gestion". Il est donc possible que l'influence de l'âge sur la productivité et donc sur le salaire ne soit pas linéaire, mais suive plutôt une courbe en forme de cloche. Ceci suggère qu'il faudrait peut-être introduire également l'âge au carré dans les équations. Nous ne l'avons cependant pas fait pour éviter de multiplier le nombre, déjà fort grand, des variables explicatives.

L'état civil a été réduit à "marié" ou "non marié". Le fait d'être marié peut être interprété de manière différente par l'employeur selon le sexe de l'individu. En effet, pour les hommes le mariage peut être vu comme un indicateur de stabilité (un homme marié ne quittera généralement pas son travail sans de bonnes raisons, car il a des bouches à nourrir) ou de responsabilité (il sera plus responsable, puisqu'il doit assumer sa famille). Pour ce qui est des femmes, être mariée signifie avoir d'autres charges importantes (s'occuper du ménage, des enfants), ce qui peut diminuer sa disponibilité (par exemple, heures supplémentaires difficiles ou impossi-

bles à accepter). "**Marie**" est une variable binaire qui vaut 1 si la personne est mariée et 0 sinon.

En introduisant la nationalité comme variable explicative, on soulève la possibilité d'une autre forme de discrimination: à productivité égale, les étrangers sont-ils moins payés que les Suisses ? Par ailleurs, la nationalité peut aussi être interprétée comme une présomption de stabilité plus grande ou plus petite. Cette forme de discrimination possible est aussi intéressante que celle que nous étudions, mais nous ne pousserons cependant pas plus loin dans cette voie. Pour cette variable, 0 = étranger, 1 = Suisse.

Le sexe de l'individu est bien entendu introduit dans les équations afin d'identifier et de mesurer une éventuelle discrimination salariale selon le sexe. La variable muette **Sexe** prend la valeur 1 si la personne interrogée est une femme et 0 si c'est un homme.

Le niveau d'éducation de l'employé est certainement une variable qui détermine le niveau des salaires : plus la personne a une formation poussée, plus grande sera sa productivité probable. Cette variable est codée comme suit :

- 1 : école obligatoire ou formation élémentaire
- 2 : apprentissage, école professionnelle, école de formation générale ou maturité
- 3 : maîtrise ou école technique
- 4 : haute école professionnelle, haute école technique ou université.

Comme dit à la note 32 dans le texte principal, cette variable soulève un problème : une formation acquise dans une haute école signifie-t-elle nécessairement, en moyenne, une productivité quatre fois plus élevée que celle associée à une formation qui s'est arrêtée au niveau de l'école obligatoire ?

2. Variables concernant l'individu et l'entreprise

Le taux d'occupation est le rapport entre le temps travaillé par l'employé et le temps de travail moyen considéré comme normal en Suisse au moment de l'enquête (40 h). Ce taux est évidemment un facteur qui influence directement les salaires. On peut cependant se demander si le taux d'occupation a une influence linéaire sur la productivité : un employé donné qui ne travaillerait que 20 heures, par exemple, produira en général plus que la moitié de ce qu'il produirait s'il travaillait 40 heures ; en sens inverse, il pourrait aussi s'investir moins dans un emploi à temps partiel que dans un emploi à temps complet.

Le type de contrat indique si le contrat est limité dans le temps (**contrat** = 1) ou non (**contrat** = 2). Avoir un contrat limité dans le temps peut réduire le niveau du salaire. En effet, lorsque le contrat n'a pas de limite prédéterminée, le travailleur sera peut-être incité à s'investir plus pleinement dans l'entreprise.

L'ancienneté est mesurée par le temps que l'employé a passé dans l'entreprise. On considère en général qu'une meilleure connaissance de l'entreprise entraîne une plus grande productivité. L'ancienneté devrait donc avoir une influence positive sur le salaire, avec toutefois la possibilité de rendements décroissants. Afin de rendre compte de cette possibilité, il conviendrait en principe d'introduire l'ancienneté au carré. Le codage de cette variable ne le permet cependant pas puisque **ancien** prend, dans l'enquête, la valeur 0 si l'employé est depuis moins de 8 ans dans l'entreprise et sinon la valeur 1.

L'expérience générale, c'est-à-dire pas nécessairement celle dans l'entreprise où l'employé travaille au moment de l'enquête, est mesurée par le nombre d'années passées dans le marché

du travail sans longue interruption. Le salaire devrait être influencé positivement par cette variable. Un problème avec cette variable est qu'on ne tient pas compte des changements d'orientation qui peuvent se produire durant la vie professionnelle. Etant donné que la loi des rendements décroissants s'applique sans doute ici aussi, il conviendrait en principe d'introduire également l'expérience au carré, mais comme dans le cas de la variable ancienneté, cela n'est pas possible, vu la manière dont cette variable est codée dans l'enquête : **exp** prend en effet la valeur 0 si la personne a moins de 8 années d'expérience et 1 sinon.

A première vue, la position hiérarchique devrait influencer le salaire positivement. Cette variable est codifiée de manière inverse, en cela qu'elle va de 1 = cadre supérieur à 9 = ouvrier et autre emploi non qualifié. On peut cependant se demander si cette variable a bien sa place dans les équations : en principe, la qualification et les autres attributs "objectifs" de l'employé devraient non seulement influencer le salaire, mais devraient aussi déterminer le niveau hiérarchique. Il n'est donc pas sûr que ce niveau exerce un effet indépendant sur les salaires.

3. Variables concernant l'entreprise

Le "secteur" de l'entreprise permet de distinguer les emplois dans des entreprises privées des emplois dans des entreprises publiques. Selon les études existantes, il peut y avoir une différence dans la formation des salaires dans ces deux secteurs, la productivité étant peut-être moins forte dans le secteur public, lequel serait géré moins efficacement que le secteur privé. Par ailleurs, il est aussi possible que la discrimination salariale selon le sexe n'ait pas, si elle existe, la même importance dans les deux secteurs. Cette variable vaut 1 si l'entreprise est dans le secteur privé et 0 si elle est dans le secteur public.

Par "branche économique", on entend le secteur primaire, secondaire ou tertiaire. Cette variable peut avoir un impact sur le niveau des salaires dans la mesure où la rentabilité des entreprises accuse des différences significatives et durables selon la branche où elles sont actives, la présomption étant que le secteur tertiaire tendait en général à être plus profitable au moment où l'enquête a été faite. La variable **div** vaut 1 si l'entreprise est dans le secteur tertiaire et 0 sinon.

Selon Flückiger et Ramirez (2000), la taille de l'entreprise a une influence positive sur les salaires. Premièrement, une grande entreprise a en général une plus grande part de marché et peut-être aussi plus de pouvoir de marché qu'une petite entreprise. Une grande entreprise générera donc en général plus de profit, qu'elle pourra partager dans une certaine mesure avec ses employés. Deuxièmement, et en sens inverse, la grande taille de l'entreprise rend peut-être plus difficile le contrôle du travail et de la productivité des employés. Le signe attendu de l'effet net de cette variable sur les salaires est donc incertain.

A.2. Régression directe générale (hommes et femmes)

Dependent Variable: LN

Method: Least Squares

Date: 05/16/01 Time: 13:22

Sample(adjusted): 2 9922

Included observations: 3666

Excluded observations: 6255 after adjusting endpoints

White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.799246	0.134261	58.09008	0.0000
AGE	0.004054	0.001589	2.552216	0.0107
EDU	0.195326	0.013499	14.47012	0.0000
SEXE10	-0.160109	0.028694	-5.579929	0.0000
MARIE	0.033908	0.022410	1.513059	0.1304
NAT	0.097632	0.022044	4.429009	0.0000
TAUXOCCUP	0.021859	0.000900	24.28785	0.0000
CONTRAT	0.198110	0.055731	3.554735	0.0004
ANCIEN	0.035037	0.021665	1.617195	0.1059
TAILETA	0.001836	0.000260	7.063615	0.0000
DIV	0.067074	0.015155	4.425766	0.0000
EXP	0.152846	0.029252	5.225083	0.0000
HIERAR	0.004115	0.003614	1.138779	0.2549
SECTEUR	0.025269	0.026890	0.939711	0.3474
R-squared	0.532374	Mean dependent var		10.84725
Adjusted R-squared	0.530709	S.D. dependent var		0.815780
S.E. of regression	0.558849	Akaike info criterion		1.677936
Sum squared resid	1140.563	Schwarz criterion		1.701639
Log likelihood	-3061.656	F-statistic		319.8197
		Prob(F-statistic)		0.000000

LN : logarithme naturel du salaire annuel brut

Cette première régression – assez bonne, puisqu'elle explique 53 % de la variation du logarithme naturel du salaire – indique qu'il y a discrimination significative à l'encontre des femmes.

A.3. Régressions directes séparées pour les femmes et les hommes

Pour les femmes uniquement :

Dependent Variable: LN
Method: Least Squares
Date: 05/16/01 Time: 14:07
Sample(adjusted): 2 4610
Included observations: 1669
Excluded observations: 2940 after adjusting endpoints
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.746146	0.229245	33.78983	0.0000
AGE	0.001778	0.002982	0.596227	0.5511
EDU	0.217566	0.030681	7.091236	0.0000
MARIE	-0.061106	0.053846	-1.134822	0.2566
NAT	0.060438	0.035781	1.689087	0.0914
TAUXOCCUP	0.020406	0.001360	15.00795	0.0000
CONTRAT	0.270187	0.095929	2.816522	0.0049
ANCIEN	0.062448	0.044346	1.408190	0.1593
TAILETA	0.002643	0.000521	5.074853	0.0000
DIV	0.053658	0.031271	1.715887	0.0864
EXP	0.129164	0.050360	2.564803	0.0104
HIERAR	0.010482	0.006884	1.522661	0.1280
SECTEUR	0.070606	0.054191	1.302900	0.1928
R-squared	0.424940	Mean dependent var	10.47905	
Adjusted R-squared	0.420773	S.D. dependent var	0.956391	
S.E. of regression	0.727881	Akaike info criterion	2.210399	
Sum squared resid	877.3656	Schwarz criterion	2.252616	
Log likelihood	-1831.578	F-statistic	101.9749	
		Prob(F-statistic)	0.000000	

Pour les hommes uniquement :

Dependent Variable: LN
Method: Least Squares
Date: 05/16/01 Time: 14:22
Sample(adjusted): 2 5312
Included observations: 1997
Excluded observations: 3314 after adjusting endpoints
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.790125	0.163186	47.73783	0.0000
AGE	0.005833	0.001116	5.225840	0.0000
EDU	0.177932	0.010352	17.18855	0.0000
MARIE	0.087847	0.017266	5.087885	0.0000
NAT	0.128070	0.028123	4.553935	0.0000
TAUXOCCUP	0.022352	0.001570	14.23952	0.0000
CONTRAT	0.130086	0.057524	2.261439	0.0238
ANCIEN	0.012899	0.018940	0.681045	0.4959
TAILETA	0.001189	0.000203	5.854848	0.0000
DIV	0.072414	0.016655	4.348006	0.0000
EXP	0.160160	0.024100	6.645770	0.0000
HIERAR	-0.001002	0.003477	-0.288004	0.7734
SECTEUR	-0.011374	0.018958	-0.599963	0.5486
R-squared	0.497843	Mean dependent var	11.15498	
Adjusted R-squared	0.494805	S.D. dependent var	0.499489	
S.E. of regression	0.355022	Akaike info criterion	0.773213	
Sum squared resid	250.0644	Schwarz criterion	0.809664	
Log likelihood	-759.0537	F-statistic	163.9127	
		Prob(F-statistic)	0.000000	

Méthode d'Oaxaca-Blinder :

	Df	%Df	Ef	%Ef
Différence totale	0.223	0.291	0.544	0.709
constante	0.044			
age	0.199	0.906	-0.021	0.094
edu	-0.073	0.469	0.082	0.531
marie	0.074	0.900	0.008	0.100
nat	0.129	0.966	-0.004	0.034
tauxoccu	0.139	0.235	0.453	0.765
contrat	-0.133	0.991	0.001	0.009
ancien	-0.018	0.909	0.002	0.091
taileta	-0.063	0.913	0.006	0.087
div	0.016	0.550	-0.013	0.450
exp	0.019	0.390	0.030	0.610
hierar	-0.049	0.998	0.000	0.002
secteur	-0.062	0.999	0.000	0.001

La dissemblance des deux régressions montre que l'effet d'une variable peut être différent selon le sexe et justifie donc l'utilisation de la méthode d'Oaxaca-Blinder. Cette dernière indique l'apport explicatif des différences entre les moyennes des attributs des hommes et celles des femmes dans l'estimation du différentiel salarial moyen. Par exemple, elle nous permet de constater que 29,1% 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 (**edu**, par exemple) indiquent une discrimination ponctuelle envers les hommes. Cependant, c'est la différence totale qui permet de conclure à l'encontre de quel sexe la discrimination salariale se produit.

A.4. Régression générale inverse avec index des qualifications

Dependent Variable: Q
 Method: Least Squares
 Date: 05/16/01 Time: 13:43
 Sample(adjusted): 1377 9918
 Included observations: 3666
 Excluded observations: 4876 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.397181	0.095854	66.73903	0.0000
LN	0.407211	0.008123	50.12877	0.0000
SEXE10	-0.243822	0.013446	-18.13361	0.0000
MARIE	-0.111145	0.011973	-9.283292	0.0000
NAT	-0.044194	0.018010	-2.453815	0.0142
TAILETA	0.000330	0.000151	2.188617	0.0287
DIV	-0.052827	0.013888	-3.803782	0.0001
SECTEUR	0.010200	0.013831	0.737465	0.4609
R-squared	0.566898	Mean dependent var		10.55032
Adjusted R-squared	0.566070	S.D. dependent var		0.540921
S.E. of regression	0.356323	Akaike info criterion		0.776223
Sum squared resid	464.4425	Schwarz criterion		0.789767
Log likelihood	-1414.816	F-statistic		684.0078
		Prob(F-statistic)		0.000000

Méthode d'Oaxaca-Blinder :

	Df	%	Ef	%
Différence totale	0.268	0.369	0.458	0.631
constante	-1.194			
LN	1.122	0.715	0.448	0.285
MARIE	0.171	0.883	-0.023	0.117
NAT	0.188	0.996	0.001	0.004
TAILETA	-0.044	0.879	0.006	0.121
DIV	0.023	0.393	0.036	0.607
SECTEUR	0.001	0.105	-0.011	0.895

Dans le cas présent, nous constatons une discrimination contre les hommes. En effet, non seulement, la régression inverse avec index des qualifications débouche sur un signe négatif pour la variable **sexe10**, mais la méthode d'Oaxaca-Blinder conclut aussi que 36,9% de la différence des qualifications moyennes ne sont pas expliqués par les différences entre les moyennes des variables explicatives des hommes et celles des femmes (ici, un signe positif de **Df** signifie qu'il y a une discrimination contre les hommes, puisque dans le cas de la régression inverse, plus le coefficient est élevé et plus on demande de qualifications).

A.5. Régression générale inverse avec différents choix de la variable dépendante

Régression générale inverse avec **edu** comme variable dépendante

Dependent Variable: EDU
 Method: Least Squares
 Date: 05/16/01 Time: 13:45
 Sample(adjusted): 1377 9918
 Included observations: 3666
 Excluded observations: 4876 after adjusting endpoints
 White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.662102	0.674546	-2.464030	0.0138
LN	0.373522	0.084161	4.438180	0.0000
AGE	0.000446	0.001450	0.307322	0.7586
SEXE10	-0.247713	0.031784	-7.793690	0.0000
MARIE	0.051172	0.029456	1.737227	0.0824
NAT	0.270842	0.045949	5.894388	0.0000
TAUXOCCUP	-0.004685	0.002160	-2.169042	0.0301
CONTRAT	-0.206608	0.074065	-2.789565	0.0053
ANCIEN	-0.107586	0.030712	-3.503053	0.0005
TAILETA	0.001027	0.000341	3.013306	0.0026
DIV	0.169788	0.031465	5.396142	0.0000
EXP	-0.152023	0.037018	-4.106716	0.0000
HIERAR	-0.014737	0.005685	-2.592164	0.0096
SECTEUR	0.020562	0.030497	0.674227	0.5002
R-squared	0.156291	Mean dependent var		2.223950
Adjusted R-squared	0.153288	S.D. dependent var		0.839856
S.E. of regression	0.772810	Akaike info criterion		2.326244
Sum squared resid	2181.103	Schwarz criterion		2.349947
Log likelihood	-4250.006	F-statistic		52.03908
		Prob(F-statistic)		0.000000

Méthode d'Oaxaca-Blinder pour **edu** = variable indépendante

	Df	%	Ef	%	
Différence totale	4.738	0.716	1.877	0.284	1.000
constante	-0.658				
AGE	0.152	0.924	-0.013	0.076	1.00
LN	6.468	0.881	0.871	0.119	1.00
MARIE	-0.010	0.661	-0.005	0.339	1.00
NAT	0.007	0.016	-0.406	0.984	1.00
TAUXOCCU	-1.032	0.473	1.150	0.527	1.00
CONTRAT	0.072	0.305	0.164	0.695	1.00
ANCIEN	-0.032	0.594	0.022	0.406	1.00
TAILETA	-0.019	0.445	-0.024	0.555	1.00
DIV	-0.136	0.640	-0.076	0.360	1.00
EXP	-0.132	0.412	0.189	0.588	1.00
HIERAR	0.056	0.646	0.031	0.354	1.00
SECTEUR	0.002	0.083	-0.024	0.917	1.00

On constate ici un cas de discrimination à l'encontre de la gent masculine. En effet, à salaire égal, l'employeur exige des employés un niveau de formation supérieur à celui des employées (c'est le signe du coefficient de la variable **sexe10** qui le montre : le fait d'être un homme a un effet positif sur le niveau de formation exigé). De plus, la méthode d'Oaxaca-Blinder renforce cette conclusion, puisqu'elle montre que seuls 28,4% du différentiel des niveaux d'éducation

moyens sont expliqués par les différences dans les moyennes des variables explicatives, les 71,6% restants étant le résultat de la discrimination contre les hommes (71,6% des différences des niveaux d'éducation masculins et féminins ne sont pas expliqués et peuvent donc refléter une discrimination à l'encontre des hommes). Cependant, vu que la régression n'explique que 15,6% de la variation du niveau d'éducation, il faut rester prudent quant à l'interprétation de ces résultats.

Régression générale inverse avec **tauxoccup** comme variable dépendante

Dependent Variable: TAUXOCCUP
 Method: Least Squares
 Date: 05/16/01 Time: 13:47
 Sample(adjusted): 1377 9918
 Included observations: 3666
 Excluded observations: 4876 after adjusting endpoints
 White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-47.78728	23.94435	-1.995765	0.0460
LN	15.00157	2.392012	6.271526	0.0000
AGE	-0.282451	0.032827	-8.604167	0.0000
EDU	-1.681386	0.669821	-2.510201	0.0121
SEXE10	-8.988013	1.177608	-7.632430	0.0000
MARIE	-7.020130	0.634624	-11.06188	0.0000
NAT	-4.777592	0.684397	-6.980736	0.0000
CONTRAT	-1.000263	1.685109	-0.593590	0.5528
ANCIEN	0.685276	0.599206	1.143639	0.2528
TAILETA	0.006623	0.007038	0.941059	0.3467
DIV	-2.810947	0.459725	-6.114405	0.0000
EXP	3.707312	1.051612	3.525361	0.0004
HIERAR	-0.025415	0.100081	-0.253948	0.7996
SECTEUR	0.091704	0.611675	0.149922	0.8808
R-squared	0.538107	Mean dependent var		83.65125
Adjusted R-squared	0.536463	S.D. dependent var		21.50310
S.E. of regression	14.64008	Akaike info criterion		8.209214
Sum squared resid	782739.9	Schwarz criterion		8.232917
Log likelihood	-15033.49	F-statistic		327.2765
		Prob(F-statistic)		0.000000

Méthode d'Oaxaca-Blinder avec **tauxoccup** = variable dépendante

	Df	%	Ef	%	
Différence totale	11.736	0.258	33.804	0.742	1.000
constante	-23.373				
AGE	7.194	0.403	10.639	0.597	1.00
EDU	-1.913	0.330	3.890	0.670	1.00
MARIE	7.398	0.985	0.111	0.015	1.00
NAT	10.131	0.735	3.651	0.265	1.00
LN	11.933	0.451	14.499	0.549	1.00
CONTRAT	0.725	0.746	0.246	0.254	1.00
ANCIEN	0.358	0.693	-0.159	0.307	1.00
TAILETA	-1.470	0.800	0.368	0.200	1.00
DIV	1.705	0.482	1.832	0.518	1.00
EXP	-2.347	0.804	-0.573	0.196	1.00
HIERAR	1.310	0.736	-0.470	0.264	1.00
SECTEUR	0.084	0.267	-0.230	0.733	1.00

Selon ces résultats, les hommes sont à nouveau discriminés. Pour le même salaire, on demande en effet aux hommes de travailler plus (c'est-à-dire d'avoir un taux d'occupation plus élevé) qu'on ne le demande aux femmes. Cela est confirmé par les résultats de la méthode d'Oaxaca-Blinder, puisque 25,8% de la différence entre le taux d'occupation des hommes et de leurs collègues féminines ne sont pas expliqués et indiquent donc une discrimination à l'encontre des hommes.

Régression générale inverse avec **ancien** comme variable dépendante

Dependent Variable: ANCIEN
 Method: Least Squares
 Date: 05/16/01 Time: 13:50
 Sample(adjusted): 1377 9918
 Included observations: 3666
 Excluded observations: 4876 after adjusting endpoints
 White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.697174	0.114178	-6.106024	0.0000
LN	0.018641	0.012128	1.537014	0.1244
AGE	0.012588	0.000695	18.10318	0.0000
EDU	-0.029933	0.008519	-3.513546	0.0004
SEXE10	-0.027125	0.015896	-1.706427	0.0880
MARIE	0.040407	0.015085	2.678646	0.0074
NAT	0.022937	0.020503	1.118694	0.2633
TAUXOCCUP	0.000531	0.000453	1.173619	0.2406
CONTRAT	0.075931	0.025938	2.927371	0.0034
TAILETA	0.001063	0.000174	6.108402	0.0000
DIV	-0.031005	0.016799	-1.845670	0.0650
EXP	0.334853	0.015269	21.93063	0.0000
HIERAR	0.001648	0.003100	0.531488	0.5951
SECTEUR	0.023006	0.015735	1.462080	0.1438
R-squared	0.309119	Mean dependent var		0.397981
Adjusted R-squared	0.306659	S.D. dependent var		0.489548
S.E. of regression	0.407633	Akaike info criterion		1.046911
Sum squared resid	606.8321	Schwarz criterion		1.070614
Log likelihood	-1904.987	F-statistic		125.6925
		Prob(F-statistic)		0.000000

Méthode d'Oaxaca-Blinder pour **ancien** = variable dépendante

	Df	%	Ef	%	
Différence totale	0.050	0.038	-1.252	0.962	1.000
constante	-0.437				
AGE	0.203	0.219	-0.723	0.781	1.00
EDU	-0.047	0.388	0.075	0.612	1.00
MARIE	-0.021	0.766	-0.006	0.234	1.00
NAT	0.104	0.562	-0.081	0.438	1.00
TAUXOCCU	0.160	0.499	-0.160	0.501	1.00
CONTRAT	0.040	0.297	-0.094	0.703	1.00
LN	0.001	0.055	0.018	0.945	1.00
TAILETA	-0.006	0.117	-0.042	0.883	1.00
DIV	0.016	0.429	0.021	0.571	1.00
EXP	-0.025	0.107	-0.206	0.893	1.00
HIERAR	0.048	0.622	-0.029	0.378	1.00
SECTEUR	0.014	0.365	-0.024	0.635	1.00

Ces résultats témoignent à nouveau d'une discrimination contre la gent masculine. Ainsi, pour avoir le même niveau de salaire que les femmes, les hommes doivent avoir passé plus de temps dans l'entreprise.

Considérant le choix de la variable dépendante, sept régressions inverses sont en principe possibles. Cependant, seules les régressions ci-dessus pour les variables **edu**, **tauxoccup** et **ancien** donnent des résultats intéressants. En effet, les quatre autres régressions ne sont pas assez robustes pour qu'on puisse en tirer des conclusions.

A.6. Correction de Heckman

Correction de Heckman générale (hommes et femmes)

```
Iteration 0: log likelihood = -8549.7593 (not concave)
Iteration 1: log likelihood = -8387.5794
Iteration 2: log likelihood = -8192.8131
Iteration 3: log likelihood = -8055.1253
Iteration 4: log likelihood = -8053.4427
Iteration 5: log likelihood = -8053.4408
Iteration 6: log likelihood = -8053.4408
```

```
Heckman selection model          Number of obs   =   10868
(regression model with sample selection)  Censored obs   =    7202
                                           Uncensored obs =    3666

Log likelihood = -8053.441          Wald chi2(13)   =   3888.55
                                           Prob > chi2    =    0.0000
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	

ln						
age	.0216782	.0009396	23.071	0.000	.0198366	.0235199
edu	.0717652	.0127826	5.614	0.000	.0467118	.0968187
sexel0	-.039079	.0229921	-1.700	0.089	-.0841427	.0059846
marie	.0148621	.0216003	0.688	0.491	-.0274738	.0571979
nat	.0906323	.0226314	4.005	0.000	.0462757	.134989
tauxoccu	.0190225	.0004293	44.314	0.000	.0181811	.0198638
contrat	.0791611	.0372692	2.124	0.034	.0061148	.1522074
ancien	.0401526	.0176502	2.275	0.023	.0055589	.0747463
tailleta	.0014172	.0001884	7.521	0.000	.0010479	.0017865
div	.0964811	.0176072	5.480	0.000	.0619717	.1309905
exp	.1316505	.0198415	6.635	0.000	.0927618	.1705392
hierar	.0001835	.0033395	0.055	0.956	-.0063619	.0067289
secteur	.010714	.0173287	0.618	0.536	-.0232496	.0446776
_cons	8.333609	.0848404	98.227	0.000	8.167325	8.499894

select						
age	-.0327449	.0008932	-36.659	0.000	-.0344956	-.0309942
edu	.2457291	.0160654	15.296	0.000	.2142416	.2772167
sexel0	-.2682985	.0272349	-9.851	0.000	-.3216778	-.2149191
marie	.1797319	.0302479	5.942	0.000	.120447	.2390168
tailleme	-.0892972	.0095386	-9.362	0.000	-.1079926	-.0706019
_cons	.8462159	.065494	12.921	0.000	.71785	.9745817

/athrho	-1.725285	.0326692	-52.811	0.000	-1.789316	-1.661255
/lnsigma	-.2724426	.0151321	-18.004	0.000	-.302101	-.2427843

rho	-.9384963	.003895			-.9456883	-.930386
sigma	.7615171	.0115233			.7392634	.7844407
lambda	-.714681	.0128293			-.739826	-.689536

```
LR test of indep. eqns. (rho = 0): chi2(1) = 1057.80 Prob > chi2 = 0.0000
```

Correction de Heckman pour les femmes uniquement

```

Iteration 0:  log likelihood = -4637.6244 (not concave)
Iteration 1:  log likelihood = -4550.338
Iteration 2:  log likelihood = -4410.4649
Iteration 3:  log likelihood = -4358.7061
Iteration 4:  log likelihood = -4356.0136
Iteration 5:  log likelihood = -4356.0072
Iteration 6:  log likelihood = -4356.0072

```

```

Heckman selection model          Number of obs   =    6387
(regression model with sample selection)  Censored obs   =    4718
                                          Uncensored obs =    1669

                                          Wald chi2(12)  =   1267.50
Log likelihood = -4356.007          Prob > chi2    =    0.0000

```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	

ln						
age	.0266914	.0016476	16.200	0.000	.0234621	.0299206
edu	.0201462	.0259591	0.776	0.438	-.0307327	.0710251
marie	.0902037	.0404671	2.229	0.026	.0108897	.1695178
nat	.0976858	.0405789	2.407	0.016	.0181526	.1772189
tauxoccu	.0171085	.0006471	26.439	0.000	.0158402	.0183767
contrat	.0994264	.0626743	1.586	0.113	-.023413	.2222659
ancien	.0842868	.0315347	2.673	0.008	.02248	.1460937
tailleta	.0016343	.000328	4.982	0.000	.0009914	.0022772
div	.1083586	.0354576	3.056	0.002	.038863	.1778542
exp	.0868003	.0316013	2.747	0.006	.0248628	.1487377
hierar	.0035372	.0058156	0.608	0.543	-.0078611	.0149356
secteur	.0156386	.0299094	0.523	0.601	-.0429826	.0742599
_cons	8.444945	.1469852	57.454	0.000	8.15686	8.733031

select						
age	-.0341343	.0012107	-28.194	0.000	-.0365073	-.0317614
edu	.2547743	.0242678	10.498	0.000	.2072103	.3023384
marie	-.005278	.0395656	-0.133	0.894	-.082825	.0722691
tailleme	-.1290482	.0130911	-9.858	0.000	-.1547063	-.1033901
_cons	.7777544	.0902577	8.617	0.000	.6008526	.9546562

/athrho	-1.928523	.0483399	-39.895	0.000	-2.023267	-1.833778
/lnsigma	.0041482	.0220051	0.189	0.850	-.0389809	.0472773

rho	-.9586138	.0039184			-.9656351	-.9501944
sigma	1.004157	.0220965			.9617691	1.048413
lambda	-.9625986	.0236744			-1.009	-.9161976

LR test of indep. eqns. (rho = 0):	chi2(1) =	671.01	Prob > chi2 =	0.0000		

Correction de Heckman pour les hommes uniquement

Iteration 0: log likelihood = -3315.5958
 Iteration 1: log likelihood = -3315.4756
 Iteration 2: log likelihood = -3315.4739
 Iteration 3: log likelihood = -3315.4739

Heckman selection model	Number of obs	=	4481
(regression model with sample selection)	Censored obs	=	2484
	Uncensored obs	=	1997
	Wald chi2(12)	=	1968.93
Log likelihood = -3315.474	Prob > chi2	=	0.0000

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	

ln						
age	.0056488	.0013032	4.334	0.000	.0030945	.0082031
edu	.1790267	.0110064	16.266	0.000	.1574545	.200599
marie	.089765	.0202032	4.443	0.000	.0501675	.1293625
nat	.128117	.0237173	5.402	0.000	.081632	.174602
tauxoccu	.0223555	.0007282	30.698	0.000	.0209282	.0237828
contrat	.1300956	.040714	3.195	0.001	.0502976	.2098936
ancien	.0129186	.0188997	0.684	0.494	-.0241241	.0499613
taileta	.0011891	.0002028	5.864	0.000	.0007917	.0015865
div	.0724873	.0169397	4.279	0.000	.0392862	.1056884
exp	.1603505	.023627	6.787	0.000	.1140425	.2066585
hierar	-.0010016	.0035497	-0.282	0.778	-.0079588	.0059556
secteur	-.0114251	.0187885	-0.608	0.543	-.04825	.0253997
_cons	7.787376	.0972907	80.042	0.000	7.59669	7.978062

select						
age	-.0420396	.0015727	-26.730	0.000	-.0451221	-.0389571
edu	.2316969	.0228172	10.154	0.000	.186976	.2764179
marie	.6818658	.0558275	12.214	0.000	.572446	.7912857
tailleme	-.1536963	.0202147	-7.603	0.000	-.1933163	-.1140762
_cons	1.271893	.1034309	12.297	0.000	1.069173	1.474614

/athrho	.0223344	.1181773	0.189	0.850	-.2092888	.2539576
/lnsigma	-1.038697	.0159017	-65.320	0.000	-1.069863	-1.00753

rho	.0223307	.1181183			-.2062857	.2486352
sigma	.3539157	.0056279			.3430554	.3651198
lambda	.0079032	.0418166			-.0740558	.0898622

LR test of indep. eqns. (rho = 0):	chi2(1) =	0.03	Prob > chi2 =	0.8553		

Méthode d'Oaxaca-Blinder :

	Df	%	Ef	%	
Différence totale	-1.008	0.264	-2.810	0.736	1.000
constante	-0.658				
age	-1.034	0.789	-0.277	0.211	1.00
edu	0.291	0.471	-0.326	0.529	1.00
marie	0.000	0.004	-0.049	0.996	1.00
nat	0.058	0.200	-0.231	0.800	1.00
tauxoccu	0.375	0.190	-1.596	0.810	1.00
contrat	0.029	0.191	-0.123	0.809	1.00
ancien	-0.025	0.927	-0.002	0.073	1.00
tailleta	-0.019	0.273	-0.051	0.727	1.00
div	-0.031	0.331	-0.062	0.669	1.00
exp	0.046	0.304	-0.105	0.696	1.00
hierar	-0.019	0.819	0.004	0.181	1.00
secteur	-0.021	0.706	0.009	0.294	1.00

La correction de Heckman requiert l'introduction d'une nouvelle variable, soit la probabilité de participation au marché de travail, laquelle est déterminée par une équation distincte. Cette équation fait appel aux quatre variables explicatives suivantes.

Nous retrouvons tout d'abord la variable **marie** (état civil) qui a une influence certaine sur la décision de participer au marché du travail selon que la personne interrogée est de sexe masculin ou féminin. En effet, la vieille tradition n'est sans doute pas entièrement morte selon laquelle une femme mariée reste au foyer pour s'occuper des diverses tâches ménagères alors que l'homme marié devra assumer la charge financière de la famille.

Le niveau de formation explique aussi la participation au marché du travail. Son effet semble clair : les personnes ayant entrepris et achevé des études veulent obtenir des rendements sur cet investissement et ont donc – toutes autres choses égales – un plus grand incitant à travailler.

Une nouvelle variable explicative est la taille du ménage, c'est-à-dire le nombre de personnes dans la famille de la personne interrogée (ce qui dépend surtout du nombre d'enfants dans le ménage). Pour les hommes, cette variable a le même effet attendu que la variable **marie** : plus il y a de bouches à nourrir, plus l'homme est poussé à travailler. Du côté de la femme, c'est le contraire : s'il y a des enfants dans le ménage, la femme aura davantage tendance à rester à la maison pour s'en occuper.

Finalement, l'âge a une influence sur la probabilité qu'ont les femmes de travailler. En effet, au début de sa vie active ou peu après son début, une femme aura une probabilité de participation au marché du travail relativement faible dans la mesure où elle souhaite fonder une famille. Cependant, à un âge plus avancé (mettons dès 35 ans environ), la femme verra ses tâches familiales diminuer et voudra donc peut-être participer à nouveau au marché du travail.

Les tableaux de Heckman ci-dessus se divisent en deux parties, la première étant l'estimation de l'équation du salaire et la deuxième donnant les résultats pour l'équation de la probabilité de travailler.

La mesure *arthrho* et le test *Lr test of independant equations* permettent de justifier ou de réfuter l'utilisation de la correction de Heckman. Ces deux mesures calculent ρ [= $\text{cov}(u_{n1}^\circ, u_{n2}^\circ)$]. Si ces résultats permettent de rejeter l'hypothèse nulle que $\rho = 0$, cela signifie

qu'il y a effectivement un lien entre les deux équations et que la correction de Heckman est nécessaire sous peine de biais.

Cela dit, examinons les résultats obtenus avec la correction de Heckman.

La régression générale montre une discrimination salariale à l'encontre des femmes. Cette régression ne diffère pas de manière significative de la première régression (voir A.2) effectuée sans la correction de Heckman. Cependant, si l'on considère les régressions directes pour les femmes uniquement avec et sans la correction de Heckman (voir A.3), leur divergence est frappante. Cela permet de constater que, dans ce cas, le biais de sélection est important et il s'ensuit que la méthode de Heckman est nécessaire dans les équations pour les femmes, mais non dans celles pour les hommes (pour eux, l'hypothèse nulle $\rho = 0$ ne peut pas être rejetée). On constate d'ailleurs qu'il y a peu de différences entre la régression directe pour les hommes uniquement avec correction de Heckman et celle sans cette correction (voir A.3.), ce qui n'est pas étonnant, puisque la probabilité des hommes de participer au marché du travail tend vers l'unité.

Enfin, la méthode d'Oaxaca-Blinder permet de constater une discrimination contre les hommes alors que la régression générale montrait une discrimination contre les femmes.