

Genre et Marché du travail au Cameroun

Par

Woubou Hugues Nérée

Sous la co-direction de

Dr. E. P. Ndong Nguéma et M. Atéwamba Calvin

et sous la supervision de

Pr. Henri Gwét

Octobre 2006

Dédicace

Je dédie ce travail à toute la famille NGANKAM de fondanti, particulièrement à :

- ma grand-mère, Maman Manilang Régine
- et à mon feu oncle Papa Datchop Boniface, paix à son âme.

Puisse ce travail vous réconforter dans la peine que vous vous êtes donnés pour faire de moi un véritable homme.

Remerciements

Ce mémoire est le fruit d'un ensemble de circonstances et d'actions favorables dont je tiens à remercier les initiateurs.

Que tous mes enseignants de Master 2 de Statistique Appliquée, en particulier aux Pr. Mme Danielle Florens (Université de Paris Dauphine), Pr. Yann Le Strat (Institut de Veille Sanitaire de Paris) et Pr. Jean-François Dupuy (Université Paul Sabatier 31062 Toulouse cedex 9), trouvent ici l'expression de ma reconnaissance pour tous les enseignements de qualité que j'ai pu recevoir d'eux.

Mes remerciements vont également à mes encadreurs :

Pr Henry Gwét ;
Dr. E.P. Ndong Nguéma ;
M. Calvin Atéwamba ;
M. Barnabé Okouda ;
M. Ambroise Abanda.

Qu'ils trouvent ici ma reconnaissance pour tout le temps qu'ils ont pu consacrer pour le suivi bout à bout de ce travail malgré leurs multiples occupations.

J'exprime aussi ma gratitude à tous mes frères et toute autre connaissance qui m'ont apportés une aide quel que soit la nature pour la réalisation de ce travail, particulièrement :

Ma maman, Mne Yanmeni Bernadette et mes sœurs, Moumeni Françoise et Nkeumani N. Jeanne ;

Ma fiancée, Mlle Sylviane J. Djampou pour toute le soutien qu'elle m'a apportée ;

Mes amis : M. Claude Tadjoumessi, M. Clovis Dongmo, M. Patrice Takam, M. Guy Mahiane, M. Tchomté Sévérin, M. Pascal Enock, M. Achille Bonso, M. Achille Hyefouais, M. Olaf Kouamo, Mlle Solange Whegang Y., M. Bernard Sieledjie, M. Francis Fongang, M. Chassem, M. Nino Alfred Djiodo, M. Alain Takeu et M. Saturnin Alogo.

Je voudrais enfin faire part de mon amitié à tous mes frères, amis et camarades de promotion pour leurs encouragements répétés et dont leur nom n'est pas cité ici. Qu'ils trouvent par ces mots, mes remerciements distingués.

Table des matières

Dédicaces	i
Remerciements	ii
Notation	vii
Sigles et Abréviations	viii
Résumé	ix
Introduction	1
1 Présentation de la structure d'accueil du stage	2
1.1 Historique	2
1.2 Description des différentes structures de l'INS	2
2 Synthèse de la littérature sur l'approche Genre et Marché du travail	5
2.1 Concept du marché du travail	5
2.2 Indicateurs du marché du travail	5
2.3 Inégalité et genre sur le marché du travail	6
3 Approche théorique et Méthodologie	9
3.1 Présentation des données de l'étude	9
3.1.1 Unité d'enquête et support de collecte	9
3.1.2 Plan d'échantillonnage	10
3.2 Méthode de mesure de la discrimination	13
3.3 Méthodes statistiques	16
3.3.1 Test d'indépendance du χ^2 -Test de comparaison de proportions	16
3.3.2 Le modèle de régression probit	16
4 Résultats des analyses et des estimations	27
4.1 Analyse descriptive des données de l'étude	27
4.1.1 Caractéristique de la population	27
4.1.2 Participation à l'activité économique	30
4.1.3 Niveau du chômage	33
4.1.4 Condition d'activité	34

4.2	Résultats de la mesure de la discrimination	38
4.2.1	Bref aperçu des tests d'hypothèses classiques utilisées lors des estimations	38
4.2.2	Estimation des modèles	38
5	Conclusion	58
6	Annexe	59
	Médiagraphie	72

Liste des tableaux

2.1	Encadré 1	6
2.2	Encadré 2	8
3.1	Tableau récapitulatif des ménages enquêtés	10
4.1	Répartition de la population totale par sexe et par grands groupe d'âge	27
4.2	Taux de scolarisation de la population totale par région et par sexe	30
4.3	Répartition de la population active selon les caractéristiques géographiques	31
4.4	Répartition de la population active selon les caractéristiques socio-démographiques et par sexe	32
4.5	Taux de chômage par sexe et selon les caractéristiques géographiques et par sexe	33
4.6	Répartition de la population active occupée par Catégorie socio-professionnelle (CSP)	35
4.7	Répartition de la population active occupée selon le revenu horaire (en FCFA)	37
4.8	Estimation probit de la fonction de participation des hommes au marché du travail : modèle permettant d'estimer le vecteur ratio de mills conditionné par les hommes	40
4.9	Estimation probit de la fonction de participation des femmes au marché du travail : modèle permettant d'estimer le vecteur ratio de mills conditionné par les femmes	44
4.10	Estimation des coefficients de la fonction de gains des hommes	47
4.11	Estimation des coefficients de la fonction de gains des femmes	48
4.12	Récapitulatif des résultats des estimations : fonctions de gains	50
4.13	Estimation des coefficients de la fonction de participation des hommes au marché du travail	51
4.14	Estimation des coefficients de la fonction de participation des femmes au marché du travail	53
4.15	Récapitulatif des résultats des estimations : modèle dichotomique de participation au marché du travail	55
4.16	Variation de la probabilité de participation au marché du travail	56

Table des figures

4.1	Répartition de la population totale	28
4.2	Courbe de variation de la population totale	29
4.3	Variation du taux de scolarisation	29
4.4	Variation du taux d'activité selon les régions	31
4.5	Variation du taux de chômage selon les régions	34
4.6	Emploi par statut des travailleurs	34
4.7	Répartition de la population active occupée selon le mode de rémunération	36
4.8	Répartition de la population active occupée selon leur revenu horaire (en FCFA)	37
4.9	Pouvoir de discrimination : aire sous la courbe ROC (Hommes)	42
4.10	Courbe de la sensibilité et de la spécificité en fonction des probabilités prédites (Hommes)	42
4.11	Graphe des résidus	43
4.12	Pouvoir de discrimination : aire sous la courbe ROC (Femmes)	45
4.13	Courbe de la sensibilité et de la spécificité en fonction des probabilités prédites (Femmes)	45
4.14	Graphe des résidus	46
4.15	Graphe des résidus (Hommes)	47
4.16	Graphe des résidus (Femmes)	49
4.17	Pouvoir de discrimination : aire sous la courbe ROC (Hommes)	52
4.18	Courbe de la sensibilité et de la spécificité en fonction des probabilités prédites (Hommes)	52
4.19	Pouvoir de discrimination : aire sous la courbe ROC (Femmes)	54
4.20	Courbe de la sensibilité et de la spécificité en fonction des probabilités prédites (Femmes)	54

Notations

- v.a.r. = variable aléatoire réelle
- Ln est la fonction logarithme népérienne
- $Pr(A)$ désignera la probabilité de réalisation de l'événement A , $E(A)$, son espérance mathématique et $var(A)$, sa variance ou sa matrice de variance covariance.
- $\hat{\beta}$ est l'estimateur de β
- **MV** : Maximum de vraisemblance
- **EMV** : Estimateur du maximum de vraisemblance
- **RV** : Rapport de vraisemblance
- Φ : fonction cumulative de la loi normale centrée réduite
- $X \sim \mathcal{B}(p_i)$, X suit une loi de Bernoulli de paramètre p_i
- ϕ : densité de la loi normale centrée réduite
- f' : la fonction dérivée de la fonction f supposée dérivable
- Λ : fonction de répartition de la loi logistique
- F_X : fonction de répartition de la v.a.r. X
- x_i^t est la transposée du vecteur colonne x_i
- $\varepsilon_i \text{ iid}(0, \sigma^2)$ signifiera que ε_i est indépendant et identiquement distribué centré et de variance σ^2
- . correspond à la virgule d'un nombre décimal dans les tableaux des estimations ; ailleurs (i.e. en dehors des tableaux), le signe " , " est effectivement utilisé pour séparer la partie entière de la partie décimale.

Sigles & abréviations

- **BIT** : Bureau International du travail
- **INS** : Institut National de la Statistique
- **BUCREP** : Bureau central des recensements et des études de population
- **DEM** : Division des enquêtes auprès des ménages
- **DSRP** : Document de Stratégie de Réduction de la Pauvreté
- **ECAM II** : Deuxième Enquête Camerounaise Auprès des ménages
- **EESI** : Enquête sur l'Emploi et le Secteur Informel
- **ENS** : Sous-Direction des études et de la normalisation statistiques
- **ENSP** : Ecole Nationale Supérieure Polytechnique
- **Minpladat** : Ministère de la Planification, de la Programmation du développement et de l'Aménagement du territoire
- **MCO** : Moindre carré ordinaire
- **OMD** : Objectifs du Millénaire pour le Développement
- **RGPH** : Recensement Général de la Population et de l'Habitat
- **UI** : Unités primaires
- **ZD** : Zone de dénombrement

Résumé

Le problème fondamental de ce travail était de faire une analyse de la discrimination faite en l'encontre des femmes sur le marché du travail au Cameroun à partir des données de l'enquête sur l'emploi et le secteur informel réalisé en 2005.

Pour cela, nous cherchons à :

- . identifier les caractéristiques des individus liés au sexe qui pourraient avoir un rapport dans sa participation au marché du travail ;
- . identifié les caractéristiques des individus liés au sexe pouvant avoir une influence sur son revenu mensuel ;
- . mettre en évidence une discrimination sur le marché du travail selon le sexe.

La méthode d'analyse que nous avons adoptée était comparative. En effet, au moyen des tests d'égalité, les indicateurs calculés chez les individus de sexe féminin sont comparés à ceux des individus de sexe masculin.

A partir des statistiques comparatives, nous avons montré que :

- les femmes participent légèrement moins à l'activité économique que les hommes. Elles sont majoritaires dans les secteurs les moins valorisants. Malgré cette position défavorable, elles ont un taux de chômage faible par rapport à celui des hommes.
- Le caractère de plus en plus informel du travail des femmes a fortement modifié leur situation sur le marché du travail ces dernières années. En effet, les femmes exercent beaucoup plus dans des travaux à temps indépendants et à domicile. Cependant, ces travaux n'offrent qu'une sécurité limitée, ne permettant pas ainsi le versement des prestations et n'assurant que de faibles revenus.
- Au Cameroun, comme ailleurs dans le monde, les femmes sont de plus en plus nombreuses à rejoindre les rangs de la population active. Cette tendance ne s'accompagne malheureusement pas d'une amélioration parallèle des emplois auxquels elles ont accès. Les femmes représentent, au Cameroun, 50,4 % de la force de travail. Le taux d'offre de travail des femmes en âge de travailler est de 68,3 % contre 74,8 % chez les hommes. Cependant, elles restent concentrées dans un nombre relativement restreint de branches et d'occupations "féminines", offrant en général une rétribution et un prestige moindre, telles que les activités de vente de marchandises dans la rue.

Nous avons aussi montré qu'il existe une certaine discrimination par le revenu et par l'emploi à l'encontre des salariées au Cameroun.

Introduction

Dans la plupart des pays d'Afrique sub-saharienne, l'inégalité entre hommes et femmes sur le marché du travail est très magnifeste. Le Cameroun, pays en plein essor économique n'en n'est pas moins épargné. En effet, beaucoup de mutations se sont observées sur son marché du travail suite à la grave crise économique dont il a été victime pendant la décennie 1984-1994. Parmi les mesures prises pour le redressement de son économie, il y avait un gel de recrutements à la fonction publique. De ce fait, on a progressivement noté, un passage d'une structure formelle de l'emploi salarié public, à une structure de l'emploi fortement dominé par le secteur informel.

C'est alors dans le cadre de la mise en œuvre du Document de Stratégie de Réduction de la Pauvreté et du suivi des Objectifs du Millénaire pour le Développement, que le Gouvernement Camerounais va affiner ses options stratégiques en matière d'emploi. L'EESI vient alors à point nommé pour permettre entre autre, une nouvelle évaluation du marché du travail.

Du point où la discrimination est un phénomène présent dans différents secteurs d'activité, dans le présent travail, nous nous intéresseront à montrer son effectivité dans le marché du travail au Cameroun. En effet, si les phénomènes liés à la discrimination par genre ont beaucoup été analysés dans les pays industrialisés, ils ont été relativement négligés dans les pays en voie de développement. La conduite de cette étude nous permettra dans un premier temps à la maîtrise de la situation et de la condition d'activité suivant le sexe dans le marché du travail. Ensuite, elle devrait nous permettre d'identifier les déterminants de la discrimination par genre du marché de travail. Enfin, nous devrions être en mesure de proposer des indicateurs fiables pour aider les politiques dans la prise de décisions pour réduire le niveau de discrimination par genre dans le marché du travail pour un système économique plus efficient.

Pour mener à bout ce travail, nous l'avons subdivisé en trois chapitres. Cependant, un chapitre (le premier) de plus a été introduit pour présenter la structure d'accueil du stage, ensuite une synthèse de la revue de la littérature fera l'objet du chapitre 2 puis, l'approche théorique et la méthodologie de l'étude au chapitre 3 et enfin les résultats et leur interprétation feront l'objet du dernier chapitre de notre travail./-

PRÉSENTATION DE LA STRUCTURE D'ACCUEIL DU STAGE

Dans ce chapitre, il est question pour nous de faire un bref aperçu de l'Institut National de la Statistique, structure où nous avons effectué notre stage pendant la période allant du 22 Mai au 31 Août 2006.

1.1 Historique

L'INS est un établissement public administratif doté de la personnalité juridique et de l'autonomie financière. Créé par décret présidentiel N°2001/100 du 20 Avril 2001 et placé sous la tutelle technique du Ministère chargé de la statistique, il est constitué des biens meubles et immeubles initialement dévolus à la Direction de la Statistique et de la Comptabilité Nationale du Ministère de la planification, de la programmation du développement et de l'aménagement du territoire, ainsi que ceux des services statistiques créés au sein des divisions économiques provinciales.

L'INS, dont les organes de gestion sont le Conseil d'Administration et le Directeur Général a pour principales missions :

- d'assurer la coordination des activités du système national d'information statistique ;
- de rendre disponible les données et indicateurs statistiques nécessaires à la gestion économique et sociale ;
- d'assurer la conservation des fichiers des recensements et enquêtes réalisés par les administrations publiques et les organismes subventionnés ou contrôlés par l'Etat ;
- de favoriser le développement des sciences statistiques et les recherches économiques relevant de sa compétence, de promouvoir la formation du personnel spécialisé pour le fonctionnement du système national d'information statistique.

1.2 Description des différentes structures de l'INS

Dans l'attente de la mise en place de toutes les organes de gestion notamment le Président du Conseil d'Administration, la DSCN assume jusqu'à nos jours les missions dévolues à l'INS. De ce fait, les structures de fonctionnement de l'institut sont celles de la DSCN composé ainsi qu'il suit :

1. **La sous-direction de la Comptabilité nationale et des Synthèses Statistiques (CSS)**

Cette sous-direction s'occupe de l'élaboration des comptes nationaux et des travaux relatifs à la conjoncture.

2. **La sous-direction de la Statistique Démographique et Sociale (SDS)**

Celle-ci se charge principalement du :

- calcul du taux d'inflation des prix sur le marché ;
- calcul de la modélisation et de la prévision de l'indice.

3. **La sous-direction des Etudes et de la Normalisation Statistique (ENS)**

L'ENS s'occupe de tout ce qui est étude et normalisation statistiques. Elle centralise l'information sur les programmes et projets statistiques des administrations et organismes publics, actualise la réglementation statistique. Elle s'occupe également de la publication et de la diffusion des documents de l'Institut. Elle a en son sein trois services :

- Le service de la Normalisation Statistique ;
- Le service des Publications, de la diffusion, de la Documentation et des archives ;
- Le service des Etudes Statistiques.

Ce service où nous avons effectué notre stage est chargée :

- ★ du développement des méthodes applicables aux travaux statistiques ;
- ★ de la préparation des projets de plans de développement à moyen terme de la statistique en fonction des orientations générales arrêtées en la matière et de l'établissement des bilans d'exécutions de ces plans ;
- ★ de la préparation, en liaison avec les autres services de la Direction, des projets de programmes annuels et pluriannuels des travaux statistiques de l'ensemble des administrations et organismes publics et du suivi de leur réalisation ;
- ★ de l'analyse des éventuelles insuffisances des systèmes d'informations statistiques, notamment sur le plan de la méthodologie ;
- ★ de l'instruction des dossiers et demandes de visas statistiques.

4. **La sous-direction de l'Informatique et des Banques de Données (IBD)**

L'IBD s'occupe du codage, du traitement des données (recueillies pendant une enquête statistique), puis de leur stockage.

5. **La sous-direction des Statistiques d'Entreprises (SE)**

Elle est chargée du traitement de la statistique en provenance des entreprises. Les données qui y sont traitées sont en général recueillies par enquête auprès de ces entreprises.

6. **La Division des Enquêtes auprès des Ménages (DEM)**

La DEM est globalement chargée de l'étude des conditions de vie de la population.

Rappelons à ce titre que notre sujet est l'un des thèmes retenus à la première phase de l'EESI menée par la DEM. Les autres thèmes sont les suivant :

- (a) Les caractéristiques et les déterminants du sous emploi ;
- (b) L'analyse de la mobilité professionnelle ;

(c) La formation et insertion des jeunes sur le marché du travail ;

(d) Les grands indicateurs de l'emploi.

7. Le secrétariat permanent du plan comptable

Il s'occupe de la comptabilité privée, diffuse les normes de comptabilité dans les entreprises privées et mène des études économiques et financières des entreprises.

SYNTHÈSE DE LA LITTÉRATURE SUR L'APPROCHE GENRE ET MARCHÉ DU TRAVAIL

Dans ce chapitre, nous présentons le concept du marché du travail et ses caractéristiques, puis le point de vue de certains auteurs par rapport à leur approche sur l'aspect genre en relation avec le marché du travail.

2.1 Concept du marché du travail

En Economie du travail, le marché du travail est un lieu capital. De ce fait, l'on ne saurait proprement parler du **marché du travail** sans définir le mot **travail**. C'est en effet "une composante de l'activité humaine ayant en projet la création, la production et l'entretien de ses biens et services" (cf [9]). Le marché renvoie généralement aux lieux de rencontre entre acheteurs et vendeurs. On peut donc dire que le marché du travail est un lieu de rencontre entre un "acheteur de travail" et un "vendeur de travail". En Economie du travail, on parle d'une demande et d'une offre de capacité de travailler. Cette demande provenant par définition d'un employeur et est le nombre d'emplois que l'employeur met sur le marché du travail ; l'offre de travail, cependant correspond au volume ou au nombre d'heures de travail que les travailleurs disposent à offrir aux employeurs. Le travail est donc un bien que détient le travailleur et le marché du travail, un lieu où la concurrence entre l'offre de travail et la demande de travail se font ressentir. Cette concurrence est généralement de deux types. Une concurrence dite parfaite (ici, les employeurs et travailleurs négocient le taux de salaire) et la concurrence imparfaite (dans ce cas, soit le travailleur, soit l'employeur a le dessus sur l'autre et par conséquent, le taux de salaire du travailleur est à négocier).

2.2 Indicateurs du marché du travail

Un indicateur permet de mesurer dans le temps et dans l'espace les progrès réalisés vers la réussite d'un objectif. On distingue généralement quatre types d'indicateurs :

- Les indicateurs de moyens ou d'intrants qui mesurent les ressources fournies par les pouvoirs publics et les bailleurs de fond. A titre d'exemple, nous avons les dépenses pour l'éducation ;

- les indicateurs de résultats opérationnels ou de réalisations directes permettant d'évaluer les résultats opérationnels générés à cette sortie (Exemple : nombre d'école construites);
- les indicateurs de résultats évaluant les résultats finaux en fonction de l'objectif poursuivi (taux de scolarisation);
- les indicateurs d'impacts qui mesurent les conséquences de l'évolution des résultats dans les dimensions clés du bien-être des populations (taux d'alphabétisation).

Afin de mieux rendre compte des nouvelles tendances de l'emploi sur le marché du travail, le BIT a mis sur pieds un certain nombre d'indicateurs appelés indicateurs clés du marché du travail (cf [2]) dont voici quelques uns :

- population en âge de travailler ;
- population active ;
- emploi ;
- population inactive ;
- horaire de travail ;
- activité habituelle ;
- situation dans la profession ;
- revenu d'activité.

Il revient à chaque pays de voir comment mettre en valeur ces indicateurs pour mieux apprécier son marché du travail. L'encadré ci-dessous donne les différents indicateurs du marché du travail au Cameroun en 1996 et en 2001.

TAB. 2.1 – Encadré 1

chômages des jeunes enfants de 15 à 24 ans
ration emploi-population
situation dans la profession
emploi par secteur d'activité
emploi dans l'économie informelle
taux d'activité
taux de chômage
taux d'inactivité
incidence de la pauvreté

2.3 Inégalité et genre sur le marché du travail

En compilant à partir de divers recueils statistiques existants, les indicateurs ci-dessus évoqués, les données montrent que partout dans le monde (cf [2]), il existe une nette disparité entre les hommes et les femmes sur le marché du travail. A titre d'exemple, d'après le BIT, "les femmes travaillent dans des secteurs différents, la durée de leur travail rémunérée est moindre, leur taux de scolarisation et d'alphabétisation sont plus bas que ceux des hommes, elles font moins souvent partie des travailleurs indépendants et sont davantage exposées au risque d'être en chômage, en situation de sous-emploi ou exclues de la population active".

J.P. Lachaud (cf [1]) confirme ces disparités à travers une études qu'il a menées dans cinq capitales d'Afrique Subsaharienne (Abidjan, Bamako, Conakry, Ouagadougou et Yaoundé). Dans cette étude, il montre comment identifier les processus de discrimination à partir de l'analyse des écarts de revenus selon le genre et par ailleurs que le plus souvent une faible participation des femmes à l'activité économique.

Les auteurs tels que A. Kouamé et D. Taméko justifient cette faible participation des femmes dans l'activité économique par le fait que beaucoup parmi elles choisissent d'abord de jouer leurs rôles de mère et d'épouse en sacrifiant ainsi leur vie professionnelle. Cette inégalité s'explique par l'accès et la participation, puis l'occupation et la rémunération. On constatera le plus souvent un taux de participation à l'activité économique plus faible chez les femmes que chez les hommes dans le secteur moderne ou formel ¹, d'où leur orientation vers le secteur informel afin de se procurer d'un revenu (cf [4]). Ils ont également montré que la discrimination dont subissent les femmes sur le marché du travail est fonction de la nature des postes qu'elles occupent, elles ont une position défavorable par rapport aux hommes. Rappelons à ce titre qu'on distingue trois formes de discrimination sur le marché du travail, chacune entraînant une compensation différente pour des individus de productivité identique (cf [4]). La première forme consiste à refuser l'accès à certains emplois à un groupe de travailleurs du fait entre autre de leur sexe ou de leur nationalité. La deuxième forme quant à elle consiste à verser à certains employés une rémunération inférieure à celle des autres effectuant les mêmes tâches. Enfin, la dernière forme exige de certains employés des qualifications supérieures à celles des autres afin qu'ils puissent bénéficier de la même rémunération qu'eux. Ces trois formes de discrimination se regroupent en deux classes. La première forme est la discrimination par l'emploi et les deux autres, la discrimination par le revenu ou discrimination salariale. Les éléments de définition abordés ici reposent uniquement sur la définition économique de la notion de discrimination. Ainsi, Stiglitz (1973) explique qu'« il y a discrimination salariale si des individus avec les mêmes caractéristiques économiques reçoivent des salaires différents et que les différences sont systématiquement corrélées avec certaines caractéristiques non économiques des individus. »

Pour leur part, Aigner et Cain (1977) précisent que « la discrimination économique existe quand les travailleurs ne reçoivent pas un salaire ou une rémunération proportionnelle à leur productivité quand, en fait, une productivité égale n'est pas récompensée par un salaire égal. », considérant alors

la discrimination économique comme étant uniquement de la discrimination salariale.

La discrimination salariale est ainsi fondée sur la valorisation de certaines caractéristiques non économiques, cest-à-dire sans lien avec la productivité de l'individu, comme son origine ou son sexe, dans le niveau de salaire. « Le fait que différents groupes de travailleurs, qu'ils soient qualifiés ou non, noirs ou blancs, hommes ou femmes, reçoivent des salaires différents, invite à l'explication que les différents groupes doivent se différencier selon certaines caractéristiques évaluées sur le marché » (Arrow, 1973). Le marché valorise donc certaines caractéristiques non économiques, non liées à la productivité de l'individu, dans la détermination de son salaire. « (...) les caractéristiques personnelles du travailleur sans lien avec la productivité sont ainsi évaluées sur le marché » (Arrow, 1973).

Damico (1987) va plus loin en désignant ces caractéristiques personnelles comme superficielles, dans le sens où celles-ci sont sans lien avec les talents effectifs ou potentiels des individus, leurs compétences ou leur conduite. En situation de discrimination salariale, la productivité

¹66.9 pour cent contre 77.5 chez les hommes au Cameroun en 2001, ECAM II-INS ; 59.5 pour cent contre 81.2 pour cent chez les hommes à Abidjan (Côte d'Ivoire), 1987

marginale ne s'égalise donc pas au niveau de salaire du travailleur. Au chapitre 3, nous présentons une méthode de mesure de la discrimination suivant ces deux classes.

Une autre étude menée par Aynaoul (cf [12]) montre que la décision de participation au marché du travail peut être formalisée par une structure à choix discret où un individu choisi de participer ou non au marché du travail. Si cette technique paraissait convainquante, la question que l'on pourrait se poser est de savoir quelles sont les variables clés susceptibles de mieux rendre compte de la fonction de participation d'un individu au marché du travail. D'après les résultats portant sur l'ECAM II, la fonction de participation au marché du travail dépendrait de deux types de variables :

- Les variables de capital humain qui déterminent les caractéristiques propres d'un individu (Exemple : sexe, âge, niveau d'étude...)
- et les variables d'environnement qui caractérisent le contexte dans lequel cet individu se trouve (milieu de résidence, religion, région de résidence).

L'encadré 2 ci-dessous synthétise les variables susceptibles de mieux expliquer la fonction de gain et la décision de participer à une activité économique retenus lors de l'ECAM 2.

TAB. 2.2 – Encadré 2

Fonction de participation au marché du travail
niveau d'instruction mesurée par le nombre d'années d'étude avec succès
âge en années révolues
sexe
la taille du ménage
l'existence d'enfants en bas âge
l'existence de femmes au chômage dans le ménage
l'existence de femmes inactive
le milieu de résidence
le statut d'occupation du chef de ménage
Fonction de gain
niveau d'instruction
âge
expérience
expérience au carré
statut
secteur institutionnel ²

Dans la suite notre travail, il nous reviendra de voir comment prendre en compte ces facteurs lors de la sélection de nos variables d'intérêts par rapport aux données que nous disposons devant afin de voir celles qui pourraient le mieux être susceptibles d'expliquer notre modèle.

APPROCHE THÉORIQUE ET MÉTHODOLOGIE

Ce chapitre présente les données utilisées pour notre étude et les méthodes statistiques utilisées pour expliquer la discrimination que l'on observe dans le marché du travail Camerounais.

3.1 Présentation des données de l'étude

Les données de notre étude sont issues de l'enquête sur l'emploi et le secteur informel réalisée au Cameroun sur l'ensemble de son territoire national du 25 Mai 2005 au 10 Juillet 2005 (EESI 2005). Elle a été menée en deux phases et son objectif principal était d'évaluer «la situation de l'emploi » (phase 1) et les "activités économiques des ménages dans le secteur informel (phase 2). Six thèmes étaient retenus au cours de la phase 1 (enquête emploi) parmi lesquels le thème que nous abordons.

3.1.1 Unité d'enquête et support de collecte

L'unité d'enquête était un ménage ordinaire et ses différents membres résidant dans le champ de l'enquête. Cependant, les membres du corps diplomatique et leurs ménages ne faisaient pas parties du champ de l'étude, ainsi que compris les ménages collectifs ¹. Les analyses de l'emploi portent essentiellement sur les individus de 10 ans et plus.

Le support de collecte des données était constitué de deux formulaires : un formulaire ménage et un formulaire individuel.

Le formulaire ménage permettait de recenser tous les membres d'un ménage et ses caractéristiques sociodémographiques (migration et cadre de vie) et le formulaire individuel était destiné à tous les membres du ménages âgés de 10 ans et plus. Ce deuxième formulaire permettait de recueillir des informations sur leur offre de travail, leurs conditions d'activité et le mode d'insertion des individus sur le marché du travail.

En définitive, **l'unité statistique** de notre tableau des données était un individu âgé de 10 ans et plus chez qui on a prélevé un certain nombre d'informations (variables) telles que sexe, niveau d'instruction, nationalité, âge, situation matrimoniale, taille du ménage, sexe du chef de ménage, catégorie socio-professionnelle...

¹ceux-ci sont par exemple, les élèves à l'internat, les militaires en caserne, les malades de longue durée admis dans les hôpitaux, les religieux, les communautés...

3.1.2 Plan d'échantillonnage

Stratification

Pour une bonne représentativité de l'échantillon choisi et pour besoin d'obtenir des résultats extrapolables à l'ensemble du territoire du Cameroun. Le champ de l'étude était divisé en 12 régions. Ces régions étaient constituées des 10 provinces et des deux grandes villes (Yaoundé et Douala). (chacune subdivisée en 3 strates), les villes de Yaoundé et de Douala étaient chacune comme une strate. On avait donc au total 32 strates. L'échantillon de chaque strate a été sélectionné à deux² degrés dans le milieu urbain (50000 habitants et plus) et à trois degrés dans les milieux semi-urbain (10000 à 49999 habitants) et rural (moins de 10000 habitants).

Base de sondage

Elle était constituée des zones de dénombrements (ZD) issues des travaux cartographiques du troisième RGPH. Le tableau 3.1 donne un récapitulatif des différentes ZD.

Tirage de l'échantillon

Dans les villes de Yaoundé et de Douala et les autres grandes villes, on a tiré, au premier degré, dans chacun de leurs six arrondissements des ZD et, au deuxième degré, des ménages. Dans les strates rurales et semi-urbaines, on tire les villes au premier, les ZD au deuxième degré et enfin les ménages au troisième degré.

On avait au total sélectionné 17800 ZD, 8540 ménages, puis 38559 individus recensés.

Le tableau ci-dessous résume ces résultats.

TAB. 3.1 – Tableau récapitulatif des ménages enquêtés

A	Milieu urbain				Milieu semi-urbain				Milieu rural					
	B	C	D	E	F	B	C	D	E	G	H	C	D	E
I	1	1547	100	1400	/	/	/	/	/	/	/	/	/	/
II	1	1561	100	1200	/	/	/	/	/	/	/	/	/	/
III	1	157	18	180	5	3	124	17	170	16	5	520	15	150
IV	2	98	17	170	12	6	206	27	270	61	6	1283	17	170
V	1	79	17	170	5	3	109	20	200	31	5	648	15	150
VI	2	361	38	380	11	6	217	23	230	44	7	2921	20	200
VII	3	261	19	190	6	3	139	17	170	23	5	298	15	150
VIII	2	218	26	260	5	3	63	17	170	18	6	1070	16	160
IX	3	313	28	280	10	5	185	17	170	31	6	1359	17	170
X	8	615	34	340	9	5	171	18	180	37	8	1368	19	190
XI	1	42	17	170	4	2	127	17	170	22	6	487	15	150
XII	4	358	25	250	3	2	57	17	170	22	6	838	16	160
XIII	29	5610	439	4990	70	38	1398	190	1900	305	60	10792	165	1650

Source : BUCREP, Travaux cartographiques du 3^{ème} RGPH, 2003

²ie en deux étapes

Légende

A= Régions d'enquêtes ;

B= Nombre de grandes villes ;

C= Nombre de ZD dans la strate ;

D= Nombre de ZD sélectionnées ;

E= Nombre de ménages enquêtés ;

F= Nombre d'arrondissements concernés (petites villes) ;

G= Nombre d'arrondissements concernés ;

H= Nombre d'arrondissements de l'échantillon ;

I= Douala ; II= Yaoundé ; III= Adamaoua ; IV= Centre-Yaoundé ; V= Est ;

VI= Extrême-nord ; VII= Littoral-Doula ; VIII= Nord ; XI= Nord-Ouest ; X= Ouest ;

XI= Sud ; XII= Sud-Ouest ; XIII= Ensemble.

Extrapolation des résultats**1. Cacul des probabilités de tirage****(a) Cas des strates Yaoundé et Douala**

Au premier degré, la probabilité de tirer une ZD k dans un arrondissement j de la ville i est donnée par :

$$p_k^1 = \frac{z_{ij}}{Z_{ij}} \quad i = 1, 2 \text{ et } j = 1 \dots, 6; \quad (3.1)$$

où z_{ij} est le nombre de ZD tirés dans l'arrondissement j de la ville i et Z_{ij} , le nombre total de ZD dans l'arrondissement j de la ville i .

Au deuxième degré, la probabilité de tirer ménage m dans une ZD de l'arrondissement j de la ville i est définie par :

$$p_m^2 = \frac{e_{ijk}}{E_{ijk}} \quad i = 1, 2 \text{ et } j = 1 \dots, 6; \quad (3.2)$$

où e_{ijk} représente le nombre de ménages tirés et à enquêter dans la ZD k de l'arrondissement j de la ville i et E_{ijk} , le nombre total de ménages dénombrés dans la ZD k de l'arrondissement j de la ville i .

On déduit alors la probabilité de tirage d'un ménage m dans une ZD k de l'arrondissement j de la ville i :

$$p_{1m} = p_k^1 \times p_m^2 = \frac{z_{ij}e_{ijk}}{Z_{ij}E_{ijk}}. \quad (3.3)$$

(b) Cas des autres strates urbaines (les grandes villes)

Au premier degré, la probabilité de tirer une ZD urbaine k dans la province i ($i = 1, \dots, 10$) est :

$$P_k^1 = \frac{z_i}{Z_i}; \quad (3.4)$$

où z_i désigne le nombre de ZD urbaines tirées dans la province i et Z_i , le nombre total de ZD urbaines dans la province i .

Au deuxième degré, la probabilité de tirer un ménage m dans une ZD urbaine k de la province i est :

$$P_m^2 = \frac{e_{ik}}{E_{ik}}; \quad (3.5)$$

où e_{ik} est le nombre de ménages effectivement tirés et enquêtés dans la ZD urbaine k de la province i et E_{ik} , le nombre total de ménages dénombrés dans la ZD urbaine k de la province i .

La probabilité de tirage d'un ménage m dans une ZD urbaine k d'une province i est alors :

$$p_{1m} = P_k^1 \times P_m^2. \quad (3.6)$$

(c) **Cas des strates semi urbaines (les petites villes)**

Au premier degré, la probabilité de tirer une petite ville j (arrondissement) appartenant à une province i ($i = 1, \dots, 10$) est donnée par :

$$P_j^1 = \frac{m_{ij}}{M_i}; \quad (3.7)$$

m_{ij} étant le nombre de ménages dénombrés dans la petite ville j de la province i et M_i , le nombre total de ménages de toutes les petites villes dans ladite province.

Au deuxième degré, la probabilité de tirer une ZD semi-urbaine k dans la petite ville j d'une province i est :

$$P_j^2 = \frac{z_{ij}}{Z_{ij}}; \quad (3.8)$$

où z_{ij} est le nombre de ZD tirées dans la petite ville j de la province i et Z_{ij} , le nombre total de ZD dans la petite ville j de cette province.

Au troisième degré, la probabilité de tirer un ménage m dans une ZD k de la petite ville j dans la province i est :

$$P_m^3 = \frac{e_{ijk}}{E_{ijk}}; \quad (3.9)$$

avec e_{ijk} (resp E_{ijk}) le nombre de ménages effectivement tirés et enquêtés dans la ZD k de la petite ville j dans la province i (resp le nombre total de ménages dénombrés dans la ZD k de la petite ville j dans la province i).

La probabilité de tirage d'un ménage m dans une ZD k de la petite ville j dans la province i est finalement :

$$P_{3m} = P_j^1 \times P_j^2 \times P_m^3. \quad (3.10)$$

(d) **Cas des strates rurales**

Le schéma de tirage dans les strates rurales est identique à celui appliqué dans les strates semi-urbaines des provinces. Ainsi, en désignant par :

m_{ij} , le nombre de ménages dénombrés dans la zone rurale de l'arrondissement j de la province i ;

M_i , le nombre total de ménages de toutes les localités rurales dans la province i ;

z_{ij} nombre de ZD tiré dans la zone rurale de l'arrondissement j de la province i ;
 Z_{ij} , le nombre total de ZD dans la partie rurale de l'arrondissement j de la province i ;

e_{ijk} , le nombre de ménages effectivement tirés et enquêtés dans la ZD rurale k d'arrondissement j dans la province i ;

E_{ijk} , le nombre total de ménages dénombrés dans la ZD rurale k de l'arrondissement j dans la province i .

La probabilité de tirage d'un ménage m dans une ZD rurale k de l'arrondissement j dans la province i est ainsi donnée par la relation :

$$P_{4m} = \frac{m_{ij} z_{ij} e_{ijk}}{M_i Z_{ij} E_{ijk}}. \quad (3.11)$$

2. Calcul des coefficients d'extrapolation

Les coefficients d'extrapolation permettent d'obtenir des résultats sur l'ensemble de la population dont un échantillon a été enquêté. Pour une ZD donnée, ce coefficient correspond à l'inverse de la probabilité de tirage d'un ménage de ladite ZD. On peut alors déterminer les coefficients d'extrapolation pour chacune des quatre types de strates.

Notons enfin que le milieu semi-urbain a été jumelé au milieu rural lors des analyses et que sa prise en compte dans la constitution de l'échantillon répondait à un souci de représentativité.

3.2 Méthode de mesure de la discrimination

Avant d'étudier la discrimination présente dans le marché du travail du Cameroun, nous allons commencer par évaluer la décision de participation au marché du travail de chaque individu.

Pour cela, nous nous sommes proposés d'estimer séparément les décisions de participation selon le sexe. Toutefois, signalons que c'est l'hypothèse selon laquelle la discrimination par sexe est effective sur le marché du travail qui nous a motivé à estimer séparément les modèles de participation selon le sexe. Les besoins de modélisation de la fonction de participation au marché du travail ont conduit au choix du **modèle de régression probit binaire**. En effet, comme nous l'avons signalé au chapitre précédent, la décision de participer au marché du travail peut être formalisée par une structure à choix discret où l'individu i de sexe s choisit ($Y_{is} = 1$) ou non ($Y_{is} = 0$) de participer au marché du travail (**cf [5]**). Le modèle s'écrit :

$$Y_{is} = \beta_{0s} + \beta_{1s} X_{is} + \beta_{2s} F_{is} + \varepsilon_{is} \quad i = 1, \dots, N; \quad (3.12)$$

où X_{is} est un vecteur de variables de capital humain ;

F_{is} un vecteur de variables d'environnement ;

β_{js} , un vecteur de paramètres à estimer ;

$s = \text{masculin (h)} \text{ ou } \text{feminin (f)}$;

N , la taille de l'échantillon ;

et ε_{is} un terme d'erreur.

Nous avons estimé les équations (3.12) par la méthode du maximum de vraisemblance probit, elles conduisent aux équations estimées de participation des femmes et des hommes au marché du travail.

Etant donné que les individus d'un groupe donné ne constituent pas forcément un échantillon aléatoire de notre population, les coefficients estimés des équations précédentes peuvent s'avérer biaisés. Pour corriger ce biais de sélection dans les fonctions de gains, nous avons utilisé une technique économétrique développée par Heckman. Elle consiste à introduire dans les fonctions de gains une variable supplémentaire, l'inverse de Mill calculée à partir des équations réduites de participation :

$$\lambda_{is} = \frac{\phi(X_{is}^t \beta_s)}{\Phi(X_{is}^t \beta_s)} \quad i = 1, \dots, N; \quad (3.13)$$

avec ϕ la densité de la loi normale centrée et réduite, Φ sa fonction de répartition; X_{is} , vecteur dont composantes sont les variables de capital humain et d'environnement, puis $\beta_s = (\beta_{1s}, \beta_{2s})$.

Les équations de gains sont alors estimées selon le sexe par :

$$\ln W_{is} = \beta'_{0s} X_{is} + \beta'_{1s} \lambda_{is} + \varepsilon_{is}; \quad (3.14)$$

avec W_{is} le salaire de l'individu i et $E(\varepsilon_{is} | Y_{is} = 1) = 0$.

Les équations (3.14) peuvent être estimées par la méthode des MCO. Notons ici, que c'est l'effet d'échelle qui nous a amené à considérer le logarithme du revenu comme variable dépendante des fonctions de gain et non le revenu. Cela nous permet par ailleurs d'interpréter l'effet d'une variable explicative sur la variable dépendante en terme de pourcentage.

A partir de ces équations de gains, le revenu R_{is} est prédit

$$\widehat{R}_{is} = W_{is}, \quad (3.15)$$

puis inséré dans les équations structurelles de participation. En effet, il permet de prendre en compte l'influence des conditions de marché. Le modèle de participation au marché du travail s'écrit alors définitivement comme suit :

$$Y_{is} = \beta_{0s} + \beta_{1s} X_{is} + \beta_{2s} F_{is} + \beta_{3s} R_{is} + \varepsilon_{is} \quad i = 1, \dots, N. \quad (3.16)$$

La technique utilisée pour la validation du modèle repose sur les tests de l'hypothèse d'égalité des coefficients des équations obtenues de l'analyse des résidus.

Une fois ces équations estimées, nous pouvons à présent chercher comment expliquer les différences de revenu constaté entre les hommes et les femmes sur le marché du travail. A ce niveau, nous nous sommes beaucoup basés sur une technique développée par Oaxaca.

Celle-ci est basée sur l'estimation de l'écart de revenu entre deux groupes disposant de caractéristiques productives moyennes et des fonctions reliant ces caractéristiques à leurs revenus différents. On compare alors leurs revenus réels moyens avec ceux qu'ils auraient obtenus s'ils étaient tous payés selon la même fonction de gain.

En effet, considérons deux groupes d'individus (1) et (2) de revenus moyens resp R_1 et R_2 . D'après la relation (3.14), R_1 et R_2 sont déterminés par des fonctions de gains respectives f_1 et f_2 :

$$\begin{aligned} R_1 &= f_1(X_1) \\ R_2 &= f_2(X_2), \end{aligned} \quad (3.17)$$

avec X_i ($i = 1, 2$), un vecteur de caractéristiques moyennes applicables au groupe i .

La différence brute entre les revenus moyens de ces deux groupes est donc :

$$\Delta R = |R_1 - R_2|, \quad (3.18)$$

Sans nuire à la généralité, supposons $R_1 > R_2$, alors $\Delta R = R_1 - R_2$ et peut être décomposé comme suit :

$$\begin{aligned}\Delta R &= (f_1(X_1) - f_1(X_2)) + (f_1(X_2) - f_2(X_2)) \\ &= E + D,\end{aligned}$$

où

$$\begin{cases} E = f_1(X_1) - f_1(X_2) \\ D = f_1(X_2) - f_2(X_2). \end{cases}$$

$f_1(X_2)$ est interprétée comme le revenu d'un individu du groupe (2) qui posséderait les caractéristiques définis par le vecteur X_2 , E mesurerait alors la différence nette de revenu entre les individus des groupes. C'est une différence qui est donc bien expliquée puisque dépendant des caractéristiques moyennes de productivité des deux groupes. cependant, la différence de revenu défini par D n'est pas expliquée, d'autant plus que les fonctions de gain sont différentes ; de ce fait, D mesure une forme de discrimination. La question que l'on pourrait se poser est de savoir s'il s'agit d'une discrimination par le revenu ou par l'emploi.

D mesure à la fois une discrimination par le revenu et une discrimination par l'emploi. La décomposition suivante entre la différence brute des revenus moyens des deux groupes nous permet de mieux appréhender cela.

Désignons par P_{i1} la proportion d'individus du groupe 1 exerçant l'emploi i , alors P_{i1} peut s'écrire :

$$P_{i1} = g_1(X_1). \quad (3.19)$$

Dans ce cas, la probabilité pour les travailleurs du groupe 2 d'avoir l'emploi i est :

$$P_{i2} = g_1(X_2). \quad (3.20)$$

La différence brute entre les revenus moyens des deux groupes peut alors s'écrire :

$$\Delta R = |R_1 - R_2| = S_i |P_{i1}R_{i1} - P_{i2}R_{i2}| \quad (3.21)$$

En se plaçant dans la condition où $R_1 > R_2$, on a : $\Delta R = S_i (P_{i1}R_{i1} - P_{i2}R_{i2})$. Nous avons la décomposition suivante de ΔR .

$$\begin{aligned}\Delta R &= S_i (P_{i1}R_{i1} - P_{i2}R_{i2}) \\ &= S_i (P_{i1}f_{i1}(X_{i1}) - P_{i2}f_{i1}(X_{i2})) \\ &= S_i [P_{i1}f_{i1}(X_{i1}) - P_{i2}f_{i2}(X_{i2}) + P_{i2}f_{i1}(X_{i1}) - P_{i2}f_{i1}(X_{i1})] \\ &= S_i [P_{i2}(f_{i1}(X_{i1}) - f_{i2}(X_{i2}))] + S_i f_{i1}(X_{i1})(P_{i1} - P_{i2}) \\ &= S_i P_{i2}(f_{i1}(X_{i2}) - f_{i2}(X_{i2})) - S_i P_{i2}f_{i1}(X_{i2}) + S_i P_{i2}f_{i1}(X_{i1}) + S_i f_{i1}(X_{i1})(P_{i1} - P_{i2}) \\ &= S_i P_{i2}(f_{i1}(X_{i2}) - f_{i2}(X_{i2})) + S_i P_{i2}(f_{i1}(X_{i1}) - f_{i1}(X_{i2})) + S_i f_{i1}(X_{i1})(P_{i1} - g_1(X_2))\end{aligned}$$

soit en définitive :

$$\begin{aligned}\Delta R &= S_i P_{i2}(f_{i1}(X_{i2}) - f_{i2}(X_{i2})) + S_i P_{i2}(f_{i1}(X_{i1}) - f_{i1}(X_{i2})) + S_i f_{i1}(X_{i1})(P_{i1} - g_1(X_2)) \\ &\quad + S_i f_{i1}(X_{i1})(g_1(X_2) - P_{i2}),\end{aligned}$$

donc,

$$\Delta R = RD + RE + JE + JD$$

où

$$\begin{cases} RD = S_i P_{i2} (f_{i1}(X_{i2}) - f_{i2}(X_{i2})) \\ RE = S_i P_{i2} (f_{i1}(X_{i1}) - f_{i1}(X_{i2})) \\ JE = S_i f_{i1}(X_{i1}) (P_{i1} - g_1(X_2)) \\ JD = S_i f_{i1}(X_{i1}) (g_1(X_2) - P_{i2}). \end{cases}$$

La quantité RD mesure la discrimination par le revenu. En effet, elle isole les effet des différences dans la structure de rémunération des deux groupes. RE traduit la partie expliquée par des différences dans les caractéristiques productives moyennes. JE et JD traduisent quant à elles la contribution de l'emploi occupé à l'écart brut de revenu moyen entre les deux groupes.

- JE traduit les différences de profession dues à des différences de caractéristiques productives moyennes.
- JD mesure la discrimination par l'emploi. En effet, elle isole les effets de l'appartenance à un groupe sur l'obtention d'un emploi.

Cette méthode de décomposition élaborée par Oaxaca permet ainsi de séparer les composantes de la différence de revenu moyen entre deux groupes et de mettre à jour les influences respectives de la discrimination par le revenu et par l'emploi.

Etant donné que les techniques économétriques d'estimation des différents modèles ci-dessus évoqués repose sur le modèle de régression probit, nous présentons dans le paragraphe qui suit l'aspect théorique de ce modèle.

3.3 Méthodes statistiques

Dans ce paragraphe, nous présentons les techniques statistiques utilisées pour la production et l'analyse des résultats du chapitre 4. Entre autres, les tests d'indépendance de χ^2 et les tests de comparaison de proportion utilisés au cours de la phase d'analyse descriptive des données, nous présenterons également les outils théoriques sur lesquels repose le modèle probit.

3.3.1 Test d'indépendance du χ^2 -Test de comparaison de proportions

Etant donné que la partie descriptive que nous présentons au chapitre 4 porte sur la comparaison des conditions et de la situation d'activité selon le genre, nous serons confrontés à la recherche d'une éventuelle dépendance entre deux variables qualitatives et dans le cas contraire, on sera confronté à un test de proportion de ces deux variables. Plus précisément, pour ce second cas, on sera amené à voir si deux proportions (pourcentages) sont significativement différents ou pas, d'où la nécessité d'un tel test.

3.3.2 Le modèle de régression probit

3.4.2.1 Introduction

Dans cette section, il est question de présenter les outils théoriques du modèle probit que nous avons utilisé pour estimer la fonction de participation au marché du travail. Nous tenons

cependant à rappeler ici que nous nous sommes beaucoup plus appuyés sur les travaux de Hecman et de la littérature pour le choix de cette méthode de régression. Toute fois, nous signalons néanmoins que le modèle régression linéaire habituellement utilisé ne pouvait être utilisé pour notre étude, puisque ses hypothèses sont violées (la variable indépendante n'est pas continue, la variance de l'erreur n'est pas constante...).

Commençons par signaler que le modèle probit est un cas du modèle dichotomique. En effet, le modèle dichotomique est un modèle statistique dans lequel la variable expliquée ne peut prendre que deux modalités : on parle alors de variable dichotomique. Ce modèle tente d'expliquer la survenue ou la non survenue d'un événement (l'individu i participe au marché du travail pour notre cas).

Soit donné un échantillon de N individus pour lesquels on a observé un certain nombre x_i d'informations. Pour un individu i ($i = 1, \dots, N$) choisi, on observe si un certain événement s'est réalisé ou pas. On note alors y_i , la variable codée associée à cet événement. y_i est donc défini par :

$$y_i = \begin{cases} 1 & \text{si l'événement s'est réalisé pour l'individu } i \\ 0 & \text{sinon.} \end{cases}$$

Pour notre cas, on a :

$$y_i = \begin{cases} 1 & \text{si l'individu } i \text{ participe au marché du travail} \\ 0 & \text{sinon.} \end{cases}$$

Comme conséquence, $(y_i | x_i) \sim \mathcal{B}(p_i)$, où $p_i = Pr(y_i = 1 | x_i)$.

Signalons qu'on aurait pu aussi coder la variable y_i autrement. Toute fois, le choix du codage que nous avons adopté nous permet directement de définir la probabilité de survenue de l'événement comme l'espérance d'une v.a.r. codée y_i . En effet, puisque $(y_i | x_i) \sim \mathcal{B}(p_i)$, on a :

$$E(y_i) = p_i. \quad (3.22)$$

Plus généralement, l'observation binaire y_i est définie comme un indicateur des valeurs prises par une variable non observable y_i^* qui fait que $y_i = 1$ lorsque y_i^* dépasse un certain seuil s et $y_i = 0$ dans le cas contraire.

Définition 3.1.

La variable inobservable y_i^* qui varie de $-\infty$ à $+\infty$ induisant la valeur de y_i selon qu'elle dépasse le seuil s ou pas est appelée variable latente et on a :

$$y_i = 1_{\{y_i^* > s\}}.$$

où la variable y_i^* s'écrit sous la forme

$$y_i^* = x_i' \beta + \varepsilon_i \quad \text{avec } \varepsilon_i \text{ iid } (0, \sigma^2)$$

L'objectif du modèle dichotomique consiste alors en l'explication de la survenue d'un événement en fonction d'un certain nombre de caractéristiques observées sur les individus de l'échantillon. L'approche généralement adoptée est de spécifier la probabilité de survenue p_i de cet événement conditionnellement aux variables explicatives $x_i = (x_i^1, \dots, x_i^p)$, $i = 1, \dots, N$. On

a alors :

$$\begin{aligned}
 p_i &= \Pr(y_i = 1 \mid x_i) \\
 &= \Pr(y_i^* > s \mid x_i) \\
 &= \Pr(x_i' \beta + \varepsilon_i > s) \\
 &= 1 - \Pr(\varepsilon_i \leq s - x_i' \beta) \\
 &= 1 - F_{\varepsilon_i} \left(\frac{s}{\sigma} - x_i' \frac{\beta}{\sigma} \right)
 \end{aligned}$$

c'est-à-dire,

$$p_i = 1 - F_{\varepsilon_i} \left(\frac{s}{\sigma} - x_i' \frac{\beta}{\sigma} \right). \quad (3.23)$$

Un choix intuitif pour cette modélisation est d'utiliser une fonction de répartition ou une fonction cumulative de densité de l'erreur ε_i . Ce choix est basé sur les fonctions de répartition de la loi normale et de la loi logistique. Les raisons étant portées entre autre sur l'existence d'une table de la lois normale et de la simplicité de la loi logistique.

Définition 3.2.

Dans le cas où ε_i est distribué suivant une loi logistique, on obtient le modèle logit

$$F_{\varepsilon_i}(t) = \Lambda(t) = \frac{\exp(t)}{1 + \exp(t)} \quad (3.24)$$

et si ε_i est distribué selon une loi normale standard, on retrouve un modèle probit :

$$F_{\varepsilon_i}(t) = \Phi(t) = \int_{-\infty}^t \phi(\xi) d\xi. \quad (3.25)$$

Notre étude telle que nous avons signalé un peu plus haut portera exclusivement sur ce second cas. De l'équation (3.23) et des conditions précédentes, on a $p_i = \Phi(x_i' \beta - s)$.

Sans nuire à la généralité, on supposera dans la suite $s = 0$ et le modèle s'écrit alors :

$$p_i = \Phi(x_i' \beta) \quad (3.26)$$

Les résultats de l'analyse effectuée sous le logiciel Stata nous permet d'écrire notre modèle sous la forme :

$$\Phi^{-1}(p_i) = x_i' \beta = \sum_{j=1}^p \beta_j x_i^j \quad (3.27)$$

éventuellement augmenté du terme d'erreur.

Les techniques économétriques montrent que la spécification de la probabilité repose sur l'estimation des paramètres β_j et sous l'hypothèse d'indépendance des observations. Compte tenu du fait de la connaissance de la loi des perturbations ($\varepsilon_i \sim N(0, 1)$), nous utiliserons à cet effet la méthode du MV.

3.4.2.2 Estimation par la méthode du maximum de vraisemblance

A l'événement $y_i = 1$, (resp $y_i = 0$) est associée la probabilité $p_i = \Phi(x_i^t \beta)$ (resp $1 - p_i = 1 - \Phi(x_i^t \beta)$). Cela permet de considérer les valeurs observées y_i comme les réalisations d'un processus binomial avec une probabilité $\Phi(x_i^t \beta)$. La vraisemblance des échantillons associées à un modèle probit s'écrira donc comme la vraisemblance des échantillons associées aux modèles binomiaux. Donc, la vraisemblance associée à l'observation y_i est :

$$L(y_i, \beta) = p_i^{y_i} (1 - p_i)^{1 - y_i}.$$

Pour un échantillon de taille $N : y = (y_1, \dots, y_N)$, elle vaut :

$$L(y, \beta) = \prod_{i=1}^N L(y_i, \beta) \tag{3.28}$$

On déduit alors sa log-vraisemblance :

$$\mathcal{L}(\beta) = L(y, \beta) = \sum_{i=1}^N \{y_i \log \Phi(x_i^t \beta) + (1 - y_i) \log (1 - \Phi(x_i^t \beta))\},$$

c'est-à-dire,

$$\mathcal{L}(\beta) = \sum_{i=1:y_i=1} \log \Phi(x_i^t \beta) + \sum_{i=1:y_i=0} \log (1 - \Phi(x_i^t \beta)) \tag{3.29}$$

La méthode du MV consiste à chercher $\hat{\beta}$ tel que le système (s) suivant soit vérifié :

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \beta} = 0 \\ \frac{\partial^2 \mathcal{L}}{\partial \beta^t \partial \beta} \text{ soit } sdp. \end{cases}$$

En général, ce système n'admet pas de solution de type formel. Il est résolu grâce à des méthodes numériques parmi lesquelles la méthode du gradient, la méthode de Newton Raphson, la méthode du scoring, etc. Les estimations obtenues par ces méthodes numériques ont des propriétés désirables dont les trois plus utilisées sont données par la proposition suivante :

Proposition 3.1.

- **Convergence** : $Pr \left(\left\| \beta - \hat{\beta} \right\| > \varepsilon \right) \xrightarrow{N \rightarrow +\infty} 0 \quad \forall \varepsilon > 0.$
- **Normalité** :
 - * $\hat{\beta} \underset{N \rightarrow +\infty}{\widetilde{\mathcal{N}}} \mathcal{N} \left(\beta, var(\hat{\beta}) \right).$
 - * $\frac{\hat{\beta}_j - \beta_j}{Se \left(\hat{\beta}_j \right)} \xrightarrow{N \rightarrow +\infty} \mathcal{N} (0, 1).$
- $var(\hat{\beta})$ est asymptotiquement la plus petite parmi tous les estimateurs de β .

Pour la preuve cf [7].

Remarque 3.1.

La quantité $G(\beta) = \frac{\partial \mathcal{L}}{\partial \beta}$ de la première équation du système (s) précédent est appelé vecteur gradient au point β . Compte tenu de (3.29) On a :

$$G(\beta) = \sum_{i=1:y_i=1} \frac{\phi(x_i^t \beta)}{\Phi(x_i^t \beta)} x_i^t + \sum_{i=1:y_i=0} \frac{-\phi(x_i^t \beta)}{1 - \Phi(x_i^t \beta)} x_i^t,$$

c'est-à-dire,

$$G(\beta) = \sum_{i=1}^N \frac{(y_i - \Phi(x_i^t \beta)) \phi(x_i^t \beta)}{\Phi(x_i^t \beta) (1 - \Phi(x_i^t \beta))} x_i^t \tag{3.30}$$

Il vient alors que :

Proposition 3.2.

L'estimateur $\hat{\beta}$ du MV du vecteur de paramètre β dans le modèle probit vérifie $G(\hat{\beta}) = 0$, où G est le gradient associé à la log-vraisemblance au point β .

Comme nous l'avons dit précédemment, le système (s) ne peut être résolu que par des méthodes numériques. Cela vient du fait que sa première équation n'est pas linéaire. Les méthodes numériques évoquées plus haut utilisent toutes le gradient et la matrice des dérivées secondes définie par : $H(\beta) = \frac{\partial^2 \mathcal{L}}{\partial \beta^t \partial \beta}$ et on a en posant $x = x_i^t \beta$:

$$\begin{aligned} H(\beta) &= \frac{\partial}{\partial \beta} \left(\frac{\partial \mathcal{L}}{\partial \beta} \right)^t \\ &= \frac{\partial}{\partial \beta} G(\beta)^t \\ &= \frac{\partial}{\partial \beta} \left[\sum_{i=1}^N \frac{(y_i - \Phi(x)) \phi(x)}{\Phi(x) (1 - \Phi(x))} \right] x_i \end{aligned}$$

$$\begin{aligned} \text{i.e. } H(\beta) &= \sum_{i=1}^N \left[\frac{\phi^2(x) + (y_i - \Phi(x)) \phi'(x)}{\Phi(x) (1 - \Phi(x))} - \frac{(y_i - \Phi(x)) \phi(x) (\phi(x) (1 - \Phi(x)) - \phi(x) \Phi(x))}{\Phi^2(x) (1 - \Phi(x))^2} \right] x_i^t x_i \\ &= \sum_{i=1}^N \frac{(y_i - \Phi(x)) \phi'(x)}{\Phi(x) (1 - \Phi(x))} x_i^t x_i + \sum_{i=1}^N \frac{y_i - 2\Phi(x) y_i - \Phi^2(x)}{\Phi^2(x) (1 - \Phi(x))^2} \phi(x) x_i^t x_i \\ &= \sum_{i=1}^N \frac{(y_i - \Phi(x)) \phi'(x)}{\Phi(x) (1 - \Phi(x))} x_i^t x_i - \sum_{i=1}^N \frac{y_i \phi^2(x)}{\Phi^2(x)} x_i^t x_i + \sum_{i=1}^N \frac{(y_i - 1) \phi^2(x)}{(1 - \Phi(x))^2} x_i^t x_i \end{aligned} \tag{3.31}$$

donc

$$H(\beta) = \sum_{i=1}^N \frac{(y_i - \Phi(x_i^t \beta)) \phi'(x_i^t \beta)}{\Phi(x_i^t \beta) (1 - \Phi(x_i^t \beta))} x_i^t x_i - \sum_{i=1}^N \left[\frac{y_i}{\Phi^2(x_i^t \beta)} + \frac{1 - y_i}{(1 - \Phi(x_i^t \beta))^2} \right] \phi^2(x_i^t \beta) x_i^t x_i \tag{3.32}$$

Définition 3.3.

On appelle matrice hessienne associée à la log-vraisemblance d'un échantillon de taille N et noté $y = (y_1, \dots, y_N)$, la matrice définie par la relation (3.31).

Une autre particularité de la matrice hessienne est qu'elle simplifie le calcul de la matrice de variance covariance asymptotique de l'EMV. En effet, on a :

$$\begin{aligned} E(H(\beta)) &= E \left[\sum_{i=1}^N \frac{(y_i - \Phi(x)) \phi'(x)}{\Phi(x)(1 - \Phi(x))} x_i^t x_i - \sum_{i=1}^N \frac{y_i \phi^2(x)}{\Phi^2(x)} x_i^t x_i - \sum_{i=1}^N \frac{(1 - y_i) \phi^2(x)}{(1 - \Phi(x))^2} x_i^t x_i \right] \\ &= - \sum_{i=1}^N \left[\frac{E(y_i)}{\Phi^2(x)} + \frac{E(1 - y_i)}{(1 - \Phi(x))^2} \right] \phi^2(x) x_i^t x_i \\ &= - \sum_{i=1}^N \left[\frac{1}{\Phi^2(x)} + \frac{1}{1 - \Phi(x)} \right] \phi^2(x) x_i^t x_i, \end{aligned}$$

ie

$$E(H(\beta)) = - \sum_{i=1}^N \frac{\phi^2(x_i^t \beta)}{\Phi(x_i^t \beta)(1 - \Phi(x_i^t \beta))} x_i^t x_i \quad (3.33)$$

Remarque 3.2.

De la relation (3.33), on déduit que la matrice d'information de Fisher du paramètre β du modèle probit est :

$$I(\beta) = \sum_{i=1}^N \frac{\phi^2(x_i^t \beta)}{\Phi^2(x_i^t \beta)(1 - \Phi(x_i^t \beta))} x_i^t x_i. \quad (3.34)$$

On a alors la proposition suivante :

Proposition 3.3.

Si $I(\beta)$ est régulière, la matrice de variance covariance asymptotique de l'estimateur $\hat{\beta}$ de β est :

$$\text{var}(\hat{\beta}) = I(\beta)^{-1} = \left(\sum_{i=1}^N \frac{\phi^2(x_i^t \beta)}{\Phi^2(x_i^t \beta)(1 - \Phi(x_i^t \beta))} x_i^t x_i \Big|_{\beta=\hat{\beta}} \right)^{-1}. \quad (3.35)$$

Nous avons également les résultats donnés par la proposition suivante :

Proposition 3.4.

1. Sous les hypothèses de régularités, on a :

$$\sqrt{N} (\hat{\beta} - \beta) \xrightarrow[N \rightarrow +\infty]{Loi} \mathcal{N}(0, I(\beta)^{-1});$$

2. La fonction $\beta \mapsto \mathcal{L}(\beta)$ est strictement concave.

(Pour la preuve, cf [7], P31 et Guyon & Ndoumbé, 2005).

Remarque 3.3.

1. Le point 1) de cette proposition nous permettra de faire des tests statistiques tels que nous veront dans le paragraphe suivant.
2. Le point 2) garantit l'unicité du maximum de la fonction log-vraisemblance. En pratique, elle rassure la convergence des estimateurs du MV vers la vraie valeur β des paramètres.

3.4.2.3 Technique du choix du modèle optimal

a) Test d'hypothèse et d'adéquation du modèle

i) Test d'hypothèse

Considérons le modèle donné par la relation (3.27). Le test d'hypothèse que nous sommes amenés à effectuer consiste à vérifier si l'effet marginal du paramètre β sur la variable dépendante est nul ou non en comparant à une statistique de test calculée à l'aide des paramètres estimés ($\hat{\beta}$ et $\hat{\sigma}$) à une statistique critique. Trois types de tests sont utilisés à cet effet : soit la statistique de Wald, soit celle du multiplicateur de Lagrange, soit celle du rapport des maxima de vraisemblance.

Cas d'un test de significativité globale

On teste l'hypothèse $H_0 : \beta_0 = \beta_1 = \dots = \beta_r = 0$ contre $H_1 : \exists i = 1, \dots, r : \beta_i \neq 0$. r correspond au nombre de variables explicatives du modèle.

- La statistique du test de Wald associée à ce test est :

$$W = \hat{\beta}^t \hat{I}(\hat{\beta}) \hat{\beta}.$$

- La statistique du test de rapport de vraisemblance s'écrit :

$$LR = -2 \left[\log L(y, \hat{\beta}) - \log L(y, \hat{\beta}^c) \right],$$

où $\hat{\beta}$ et $\hat{\beta}^c$ désignent respectivement les estimateurs non contraint et contraint du vecteur β .

- Enfin, La statistique du multiplicateur de Lagrange est

$$LM = \left(\frac{\partial \log L(y, \hat{\beta}^c)}{\partial \beta^t} \right)^t \hat{I}^{-1}(\hat{\beta}^c) \left(\frac{\partial \log L(y, \hat{\beta}^c)}{\partial \beta^t} \right).$$

Remarque 3.4.

L'estimateur \hat{I} de la matrice d'information de Fisher est obtenu par la relation :

$$\hat{I} = \sum_{i=1}^N \left(\frac{\partial \log L(y, \hat{\beta}^c)}{\partial \beta^t} \right) \left(\frac{\partial \log L(y, \hat{\beta}^c)}{\partial \beta^t} \right)$$

$$\text{avec } \frac{\partial \log L(y, \hat{\beta}^c)}{\partial \beta^t} = \sum_{i=1}^N \frac{\partial \log L(y, \beta)}{\partial \beta^t}.$$

La proposition suivante donne la loi sous H_0 de chacune de ces statistiques.

Proposition 3.5.

Sous H_0 , les W, LR et LM convergent en loi quand $N \rightarrow +\infty$ vers une loi de $\chi^2(r)$.
(cf [7], P50).

Cas d'un test de significativité individuelle

Il s'agit ici de tester $H_0 : \beta_j = 0$ contre $H_1 : \beta_j \neq 0$ $j = 1, \dots, r$. Ce test permet alors de vérifier si la variable explicative x_i^j a ou non un effet sur $\Pr(y_i = 1)$.

Proposition 3.6. *Pour ce test,*

- la statistique de Wald s'écrit :

$$W_i = \frac{\widehat{\beta}_j}{se(\widehat{\beta}_j)} \xrightarrow[N \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1);$$

- celle du rapport de vraisemblance est :

$$LR_i = -2 \left[\log L(y, \widehat{\beta}_i) - \log L(y, \widehat{\beta}_i^c) \right] \xrightarrow[N \rightarrow +\infty]{\mathcal{L}} \chi^2(1);$$

- et celle du multiplicateur de lagrange est donnée par :

$$LM_i = \left(\frac{\partial \log L(y, \widehat{\beta}^c)}{\partial \beta^t} \right)^t \widehat{I}_i(\widehat{\beta}^c) \left(\frac{\partial \log L(y, \widehat{\beta}^c)}{\partial \beta^t} \right) \xrightarrow[N \rightarrow +\infty]{\mathcal{L}} \chi^2(1)$$

$$\text{avec } \widehat{I}_i(\widehat{\beta}^c) = \sum_{j=1}^N \left(\frac{\partial \log L(y_j, \beta^c)}{\partial \beta^t} \right) \left(\frac{\partial \log L(y_j, \widehat{\beta}^c)}{\partial \beta^t} \right)$$

ii) Test d'adéquation du modèle

Il est question ici d'évaluer la qualité de l'ajustement du modèle obtenu. Ce test est basé sur le choix des variables explicatives. nous présenterons dans cette sous section les critères les plus utilisés.

1) Nombre de prédiction fausses

Il permet de faire une classification des individus à partir des probabilités prédites. Le critère s'écrit sous la forme :

$$\text{nombre de prédictions fausses} = \sum_{i=1}^N (y_i - \widehat{y}_i)^2,$$

$$\text{où } \widehat{y}_i = \begin{cases} 1 & \text{si } \widehat{\Phi}(x_i^t \beta) \geq 0.5 \\ 0 & \text{si } \widehat{\Phi}(x_i^t \beta) < 0.5 \end{cases}$$

La valeur du seuil de discrimination 0.5 est programmée par défaut sous le logiciel Stata, cependant on peut soit même se fixer un seuil.

Table de classification

Observée	Prédiction		Total
	$\widehat{y}_i = 1$	$\widehat{y}_i = 0$	
$y_i = 1$	n_{11}	n_{10}	$n_{1.}$
$y_i = 0$	n_{01}	n_{00}	$n_{0.}$
Total	$n_{.1}$	$n_{.0}$	N

$$R_{cont}^2 = \frac{n_{11} + n_{00}}{N}$$

$$R_{adj}^2 = \frac{n_{11} + n_{00} - \max(n_{1.}, n_{.0})}{N - \max(n_{1.}, n_{.0})}$$

S'il y a indépendance entre les valeurs observées et les valeurs prédites, alors les effectifs dans les cellules seraient données par la relation $n_{ij} = \frac{n_{i.}n_{.j}}{N}$. R_{cont}^2 mesure le taux de réduction d'erreurs apporté par le modèle par rapport au fait d'utiliser simplement les taux marginaux.

2) Test de Hosmer et Lemeshow

Ce test permet de vérifier si le modèle choisi permet de faire un bon ajustement des données. Les étapes de ce test sont les suivantes :

Etape 1 : estimation du modèle ;

Etape 2 : évaluation des probabilités prédites : $\hat{p}_i = \widehat{\Pr}(y_i = 1 | x_i)$;

Etape 3 : ordonner les \hat{p}_i dans l'ordre croissant ;

Etape 4 : partage des \hat{p}_i en G groupes ;

Etape 5 : calcul de la statistique

si nous désignons par :

$n_g =$ nombre d'observations dans le groupe g ;

$$\hat{y}_g = \sum_i \frac{y_i}{n_g} \quad g = 1, \dots, G;$$

$$\hat{p}_g = \sum_i \frac{\hat{p}_i}{n_g};$$

alors, la statistique du test est :

$$HL = \sum_{g=1}^G \frac{(n_g \hat{y}_g - n_g \hat{p}_g)^2}{n_g \hat{p}_g (1 - \hat{p}_g)} \xrightarrow[N \rightarrow +\infty]{Loi} \chi^2 (G - 2)$$

Etape 6 : Calcul de la p - *value* :

$$p - value = \Pr (HL > \chi_{1-\alpha}^2 (G - 2))$$

Au seuil $\alpha = 0.05$,

si $p - value < 0.05$, alors le modèle n'est pas adéquat aux données de l'étude ou bien la qualité de l'ajustement est douteuse.

si $p - value \geq 0.05$, le modèle ajuste bien les données. Dès lors, on peut chercher à supprimer les individus qui s'ajustent mal (ie ont une forte influence sur la stabilité des coefficients du modèle). On va alors associer à ce test, les autres critères de sélection du bon modèle tels que le *AIC*, le *BIC*... afin de sélectionner le bon modèle.

3) Receiver operating curve

a) Sensibilité

Elle mesure la chance de classer un individu dans la catégorie $y_i = 1$ sachant qu'il soit effectivement observé dans cette catégorie.

$$sensitivity = \Pr (\hat{y}_i = 1 | y_i = 1).$$

b) Spécificité

Elle mesure le pouvoir de mal classer un individu dans la catégorie $y_i = 0$ sachant qu'il soit effectivement observé dans cette catégorie.

$$specificity = \Pr (\hat{y}_i = 0 | y_i = 0).$$

c) **Courbe ROC**

La courbe ROC est une représentation graphique de la relation existante entre la sensibilité et la spécificité d'un test, calculée pour toutes les valeurs possibles du seuil.

Définis pour des problèmes à deux classes, la courbe ROC indique la capacité à placer les positifs devant les négatifs, elle met donc en relation dans un graphique les taux de faux positifs (1-spécificité en abscisses) et les taux de vrais positifs (sensibilité en ordonnées). Pour mieux dire, elle permet la détermination et la comparaison des performances diagnostiques de plusieurs tests à l'aide de l'évaluation des aires sous la courbe. L'aire sous la courbe ROC est de ce fait, un estimateur de l'efficacité globale du test. Le critère de la qualité de l'ajustement est donné par l'aire entre la courbe ROC et la première bissectrice. Voici les critères adoptés pour la validation :

- si aire ROC ≤ 0.5 pas de discrimination : le modèle n'est pas bon
- si $0.5 < \text{aire ROC} < 0.7$ la discrimination est insuffisante
- si $0.7 \leq \text{aire ROC} < 0.8$ discrimination acceptable
- si $0.8 \leq \text{aire ROC} < 0.9$ discrimination excellente
- si $0.9 \leq \text{aire ROC} \leq 1$ discrimination exceptionnelle

4) **Autres critères d'informations**

a) **Déviante**

Elle permet de valider un sous-modèle (à mettre en index : \mathcal{M}_1 est un sous modèle de \mathcal{M}_2 si les variables explicatives de \mathcal{M}_1 sont strictement prises parmi celles de \mathcal{M}_2). Elle est basée sur deux autres critères : la maximisation du R^2 de McFadden (il mesure le pouvoir de prédiction du modèle, ou pour mieux dire, il indique la part de la variation de la variable dépendante qui est expliquée les variables explicatives) et la minimisation du critère AIC (Akaike Information criterion) :

$$R^2 \text{ de McFadden} = 1 - \frac{\log L(y, \hat{\beta})}{\log L(y, 0)}$$

$$AIC = \frac{-2 \log L(y, \hat{\beta}) + 2r}{N}$$

b) **Bayesian Information Criterion (BIC)**

On déterminera le BIC_1 et BIC_2 pour deux modèles \mathcal{M}_1 et \mathcal{M}_2 respectivement.

$$BIC = -2\mathcal{L}(\beta) - (n - r - 1) \ln N$$

Règle de décision	
Condition	Décision
$0 < BIC_1 - BIC_2 < 2$	Faible évidence de \mathcal{M}_1 par rapport à \mathcal{M}_2
$2 \leq BIC_1 - BIC_2 < 6$	positive , on choisira le modèle ayant le plus petit BIC
$6 \leq BIC_1 - BIC_2 < 10$	forte , on choisira le modèle ayant le plus petit BIC
$ BIC_1 - BIC_2 \geq 10$	très forte , on choisira le modèle ayant le plus petit BIC

5) **Analyse des résidus**

Elle est faite dans le but de vérifier s'il y a une erreur systématique et de déterminer s'il y a des observations aberrantes ou pas très bien modélisées (outliers). Nous utilisons pour ce fait,

le résidu de Pearson.

$$\begin{aligned} p_i &= E(y_i) = \widehat{\Pr}(y_i = 1 | x_i) \\ \widehat{p}_i &= \widehat{\Pr}(y_i = 1 | x_i) \\ \text{var}(y_i) &= p_i(1 - p_i) \\ \text{cov}(y_i, y_j) &= -p_i p_j \end{aligned}$$

Le résidu de Pearson est défini par :

$$r_i = \frac{y_i - \widehat{p}_i}{\sqrt{\widehat{p}_i(1 - \widehat{p}_i)}}$$

En pratique, on utilise plus le résidu standardisé de Pearson (car la variance du résidu de Pearson n'est pas une constante égale à 1) qui permet de faciliter l'interprétation. Il est donné par :

$$r_i^{std} = \frac{r_i}{\sqrt{1 - h_{ii}}},$$

où en posant $H = \widehat{\text{Var}}(x_i^t \widehat{\text{Var}} x_i^t)^{-1} x_i^t \widehat{\text{Var}}$, $h_{ii} = \widehat{p}_i(1 - \widehat{p}_i) x_i \widehat{\text{Var}}(\beta) x_i^t$, $\widehat{\text{Var}} = \text{diag}(\sqrt{\widehat{p}_i(1 - \widehat{p}_i)})$. En pratique, on s'intéresse généralement aux observations dont $|r_i^{std}| > 2$ et on utilise encore les critères *AIC*, *BIC* ou R^2 de McFadden.

Nous ne pourrions terminer ce chapitre sans toute fois présenter les techniques utilisées pour la validation des fonctions de gains. En effet, comme signaler un peu plus en arrière, la technique économétriques utilisée était basée sur la méthode des MCO. Nous savons que les modèles obtenus doivent satisfaire les conditions suivantes :

- * pas de biais systématique : $E(\varepsilon_i) = 0$;
- * il y a homoscedasticité : $\text{var}(\varepsilon_i) = \text{cste} = \sigma^2$;
- * les résidus ne doivent pas être corrélés : $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ pour $i \neq j$.
- * $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.

Etant donné que les points 1) et 2) sont inclus dans 4), en pratique, nous utiliserons les tests de

- Jarque-Bera pour vérifier le point 4 (ce test permet de vérifier si les résidus suivent une loi normale)
- White ou de Breuchg-Pagen pour le point 3 (test d'hétéroscedasticité de white).

Notons enfin pour terminer que dans le cas de comparaison de deux modèles, on utilisera le test de Fisher et la valeur de la statistique R^2 ³.

³ R^2 mesure la corrélation entre les variables explicatives (globalement) et les réponses

RÉSULTATS DES ANALYSES ET DES ESTIMATIONS

Après avoir décrit les caractéristiques productives et sociales des hommes et femmes au Cameroun en 2005, les résultats des phénomènes de discrimination sont présentés du niveau le plus fin à celui le plus vaste d'analyse de la situation professionnelle.

4.1 Analyse descriptive des données de l'étude

4.1.1 Caractéristique de la population

TAB. 4.1 – Répartition de la population totale par sexe et par grands groupe d'âge

Groupe d'âge	Sexe	
	Masculin	Féminin
0-4	50,9	49,1
5-9	52,2	47,8
10-14	49,9	50,1
15-19	50,0	50,0
20-24	45,8	54,2
25-29	47,6	52,4
30-34	48,8	51,2
35-39	46,6	53,4
40-44	45,5	54,5
45-49	46,4	53,6
55-59	56,8	43,2
60-64	44,5	55,5
65-69	45,9	54,1
70-74	56,4	43,6
75-79	56,3	43,7
80 et +	43,9	56,1
Cameroun	49,4	50,6

Source : INS, EESI 2005, Phase I

Au vu des analyses effectuées, la structure de la population montre qu'au Cameroun, il y a plus de femmes que d'hommes. Cette différence, significative au seuil de 5 pour cent, est observée à partir de l'âge des jeunes adolescents (20 ans) jusqu'à l'âge de la ménopause (54 ans). Cela pourrait s'expliquer par le fait que c'est la tranche d'âge que l'homme se fixe pour mener une vie autonome.

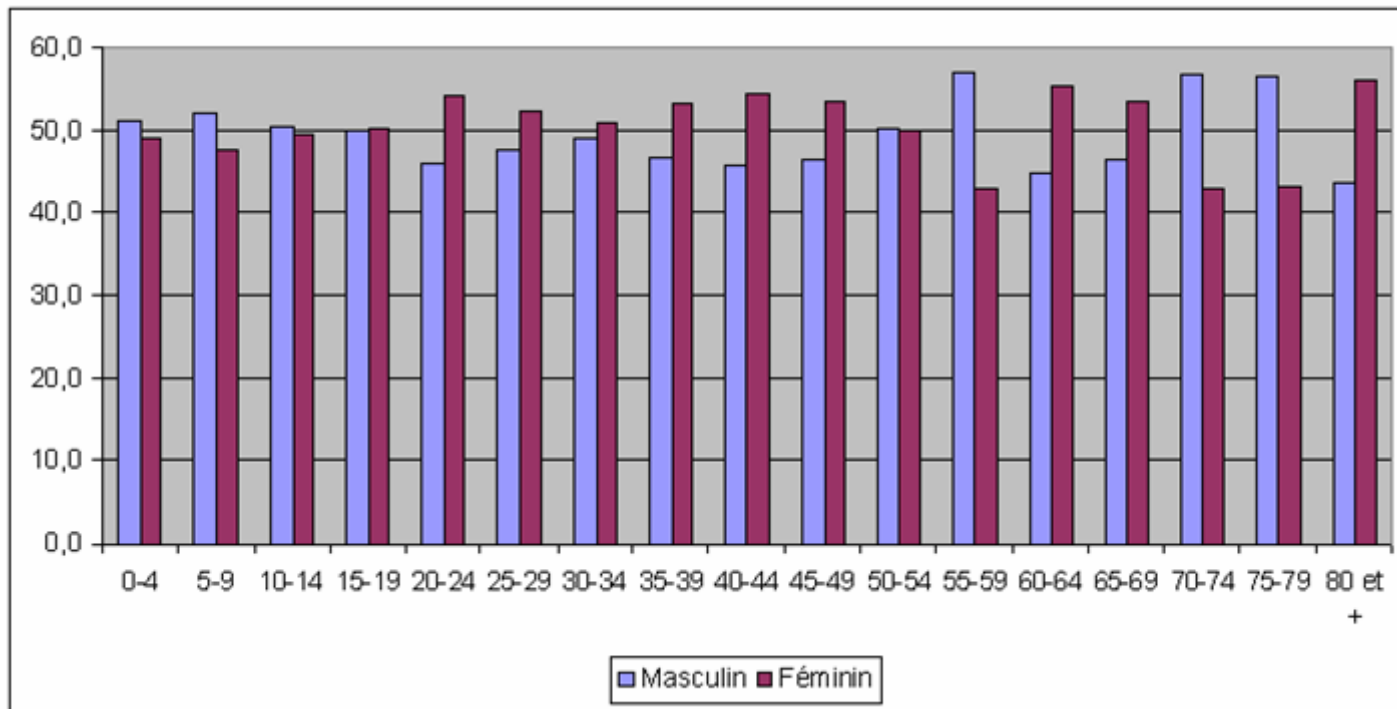


FIG. 4.1 – Répartition de la population totale

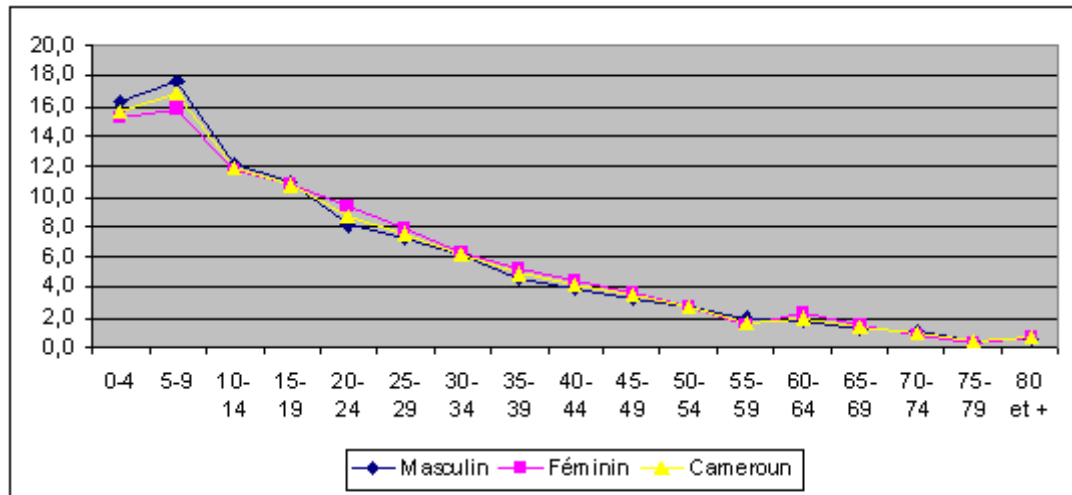


FIG. 4.2 – Courbe de variation de la population totale

Mais, curieusement on constate que quelque soit l'âge, les hommes vont davantage à l'école que les femmes. Cette tendance reste également vraie au niveau des provinces où les inégalités persistent. On note ici, un taux de scolarisation plus bas (inférieure au taux de scolarisation nationale qui est de l'ordre de 78,1 pour cent) dans la partie septentrionale du pays. C'est d'ailleurs les régions du pays où les écarts selon le genre sont les plus significatifs en matière de l'accès à l'éducation. On pourrait expliquer ce phénomène de sous scolarisation, si l'on peut dire, au poids que cette région du pays accorderais à la tradition ; cela est particulièrement visible dans les provinces du Nord et de l'Extrême-Nord.

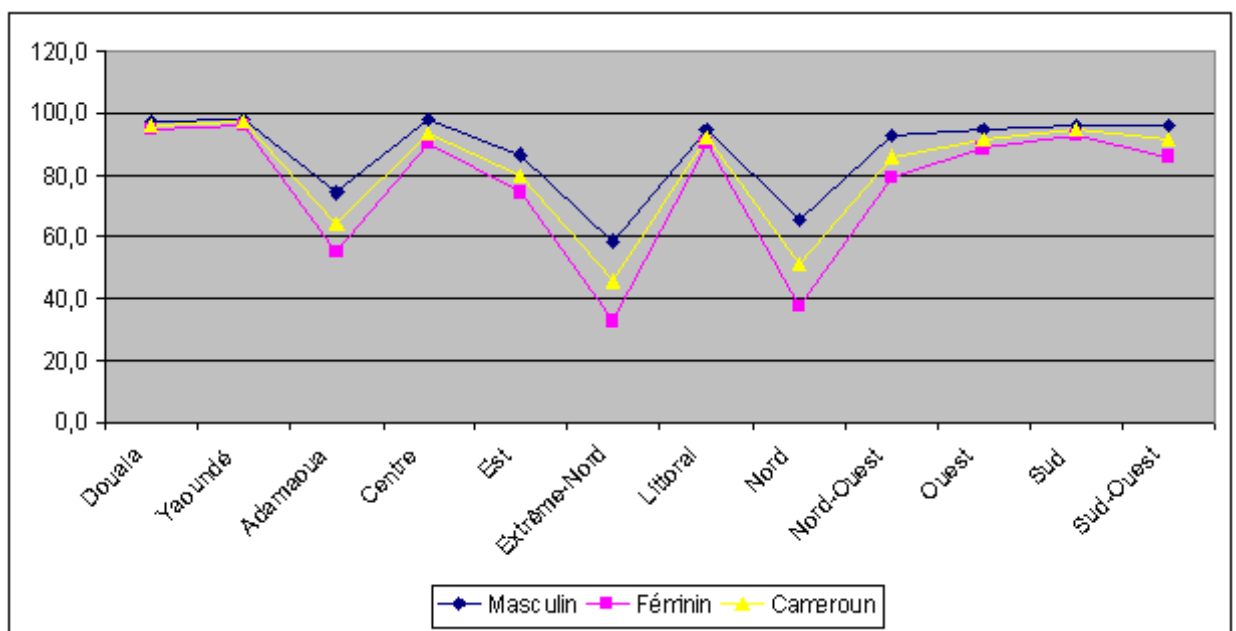


FIG. 4.3 – Variation du taux de scolarisation

TAB. 4.2 – Taux de scolarisation de la population totale par région et par sexe

	Sexe		Ensemble
	Masculin	Féminin	
Région			
Douala	97,5	94,7	96,1
Yaoundé	98,3	96,0	97,1
Adamaoua	74,6	55,1	64,3
Centre	98,0	90,0	93,8
Est	85,6	75,3	80,1
Extrême-nord	59,2	33,1	46,0
Littoral	95,0	89,7	92,3
Nord	65,5	37,7	51,4
Nord-ouest	92,7	79,5	85,7
Ouest	94,6	88,6	91,5
Sud	96,5	93,1	94,8
Sud-ouest	96,0	86,2	91,2
Milieu de résidence			
Urbain	94,5	88,9	91,7
Rural	79,1	62,8	70,7
Cameroun	84,7	71,8	78,1

Source : INS, EESI 2005, Phase I

4.1.2 Participation à l'activité économique

La participation à l'activité économique pour un individu, est le fait pour lui, d'être présent sur le marché du travail. Ainsi, sur l'ensemble de la population potentiellement active ¹, sept sur dix sont effectivement actives, ce qui correspond à un taux d'activité de 71,2 pour cent. Il est de l'ordre de 78,5 pour cent en milieu rural, contre 56,6 en milieu urbain, notons ici que cette différence est très significative au seuil de 5 pour cent. Lorsqu'on va plus loin dans les provinces, on note un taux d'activité plus élevé à l'Extrême-Nord (84,2 pour cent avec un taux de participation des hommes de l'ordre de 87,2 pour cent contre 81,4 pour les femmes ²) suivi immédiatement du Nord. On note également que seules les provinces du Nord-Ouest et de l'Ouest présentent un taux de participation à l'activité économique plus élevé chez les femmes que chez les hommes (74,3 pour cent contre 69,0 au Nord-Ouest et 74,8 pour cent contre 67,3 pour cent).

En tenant compte des caractéristiques socio-démographiques de la population, par rapport au grand groupe d'âge, on constate un faible écart de la participation à l'activité économique entre les hommes et les femmes. Cependant, selon la religion, la discrimination est beaucoup plus prononcée chez les musulmans et chez les chrétiens. Notons enfin que c'est la population non scolarisée (c'est-à-dire n'ayant jamais été à l'école, ne serait-ce qu'à l'école primaire) qui est très présent dans le marché du travail (91,0 pour cent chez les hommes contre 77,3 pour cent chez les femmes).

¹population en âge de travailler

²différence significative au seuil de 5 pour cent

TAB. 4.3 – Répartition de la population active selon les caractéristiques géographiques

	Sexe		
	Masculin	Féminin	
Région			Ensemble
Douala	69,1	51,9	60,7
Yaoundé	62,9	47,4	55,2
Adamaoua	81,6	57,1	69,0
Centre	69,5	67,0	68,2
Est	77,1	63,3	69,6
Extrême-nord	87,2	81,4	84,2
Littoral	71,7	74,5	73,2
Nord	86,7	74,7	80,6
Nord-ouest	69,0	74,3	71,9
Ouest	67,3	74,8	71,3
Sud	77,1	70,0	73,6
Sud-ouest	69,1	62,9	66,0
Milieu de résidence			
Urbain	65,8	51,4	56,6
Rural	79,9	77,1	78,5
Cameroun	74,5	68,0	71,2

Source : INS, EESI 2005, Phase I

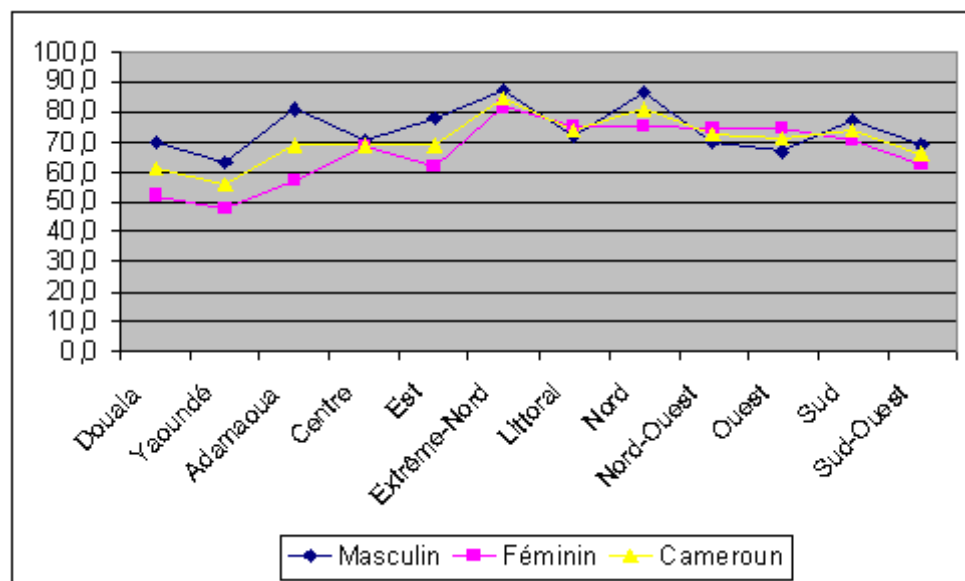


FIG. 4.4 – Variation du taux d'activité selon les régions

TAB. 4.4 – Répartition de la population active selon les caractéristiques socio-démographiques et par sexe

	Sexe	
	Masculin	Féminin
Grand groupe d'âge		
10-29 ans	62,4	57,6
30-49 ans	97,3	87,6
50 ans et plus	82,3	74,7
Réligion		
Catholique	72,7	69,3
Protestant	71,8	68,1
Autre chrétiens	80,9	65,2
Musulman	77,6	57,3
Animiste	88,8	94,7
Sans religion	84,5	77,7
autres religions	69,3	66,4
Situation matrimoniale		
Célibataire	60,1	51,3
Marié monogame	93,9	80,9
Marié polygame	94,1	81,9
veuf (Veuve)	77,5	74,1
Divorcé	90,3	86,8
Union libre	95,5	75,2
Niveau d'instruction		
Non scolarisé	91,0	77,3
Primaire	73,2	70,0
Secondaire 1 ^{er} cycle	68,7	57,4
Secondaire 2 ^{ème} cycle	71,7	56,0
Supérieur	77,4	63,2
Cameroun	74,8	68,3

Source : INS, EESI 2005, Phase I

4.1.3 Niveau du chômage

La notion du chômage abordée ici se rapporte à la définition du BIT. En effet, un individu est déclaré être en chômage s'il n'a pas travaillé ne serait-ce qu'une heure, lors de la semaine précédent l'enquête, n'a pas recherché un emploi au cours du mois précédent l'enquête, mais étant cependant disponible à travailler si on lui en offrait un emploi. Ainsi, en se référant à cette définition, le taux de chômage au Cameroun est environ 4,4 pour cent. On note cependant, un taux de chômage plus élevé chez les femmes que chez les hommes. Cela pourrait s'expliquer par le fait que la plupart des femmes qui, généralement ne travaillent pas ont plus tendance à se déclarer comme ménagères. Notons aussi que cette tendance est presque la même selon les régions. En effet, seules les provinces du littoral et du centre, présentent un taux de chômage des hommes largement supérieur à celui des femmes.

TAB. 4.5 – Taux de chômage par sexe et selon les caractéristiques géographiques et par sexe

	Sexe		
	Masculin	Féminin	
Région			Ensemble
Douala	10,6	15,1	12,5
Yaoundé	11,4	19,2	14,7
Adamaoua	1,7	1,5	1,6
Centre	6,1	3,7	4,9
Est	2,3	2,7	2,5
Extrême-nord	0,5	0,2	0,4
Littoral	5,4	4,0	4,6
Nord	0,9	3,0	1,0
Nord-ouest	1,1	2,2	3,6
Ouest	6,0	7,0	1,7
Sud	5,7	7,9	6,4
Sud-ouest	4,2	4,6	6,7
Milieu de résidence			
Urbain	8,8	13,2	10,7
Rural	1,9	1,5	1,7
Cameroun	4,2	4,6	4,4

Source : INS, EESI 2005, Phase I

Par rapport aux caractéristiques socio-démographiques, on constate que quelque soit la modalité des différentes caractéristiques retenues, le taux de chômage des hommes est supérieur à celui des femmes. Moins de femmes sont en effet, défavorisées par rapport aux hommes. Plus spécifiquement, considérant le grand groupe d'âge, le chômage touche indifféremment les hommes et les femmes âgés de 10 à 29 ans. A partir de 30 ans, ce taux est plus élevé chez les hommes que chez les femmes. Selon la religion, les taux sont sensiblement égaux entre les hommes et les femmes des religions catholiques et musulmans. Chez les protestants et les sans religions, ces taux sont plus élevés chez les hommes que chez les femmes. Par contre, pour les adeptes des autres religions, c'est la tendance inverse qui est observée. Selon le statut matrimonial, les taux sont sensiblement égaux entre hommes et femmes chez les célibataires,

largement supérieur chez les hommes et chez les veufs (ves). Par ailleurs, on constate que le chômage affecte plus les hommes non scolarisé et ceux du primaire que les femmes. A partir du secondaire, ce phénomène touche plus les femmes que les hommes.

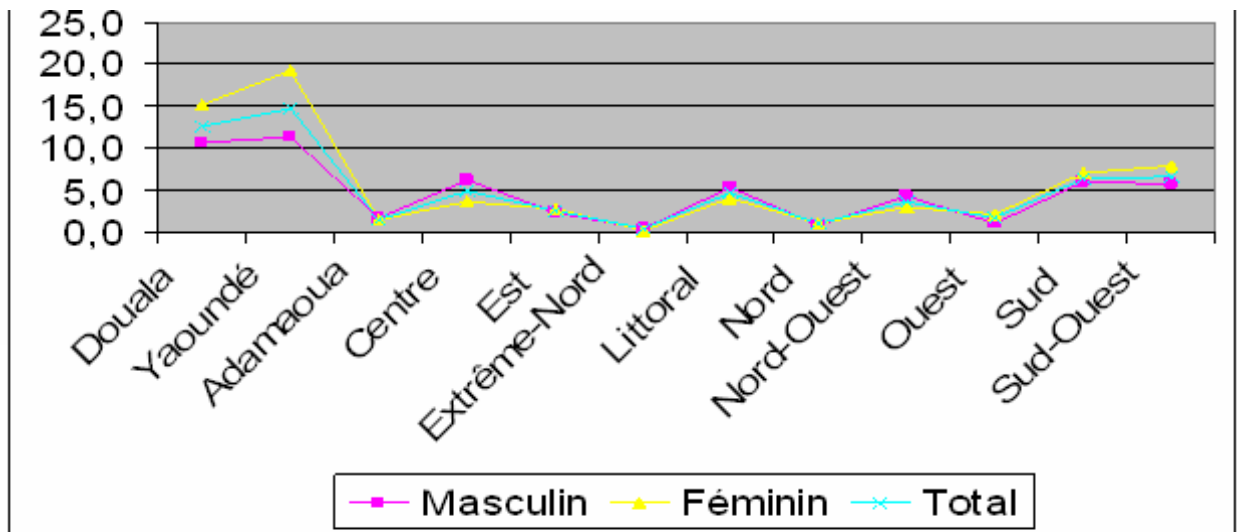


FIG. 4.5 – Variation du taux de chômage selon les régions

4.1.4 Condition d'activité

Nature des emplois occupés

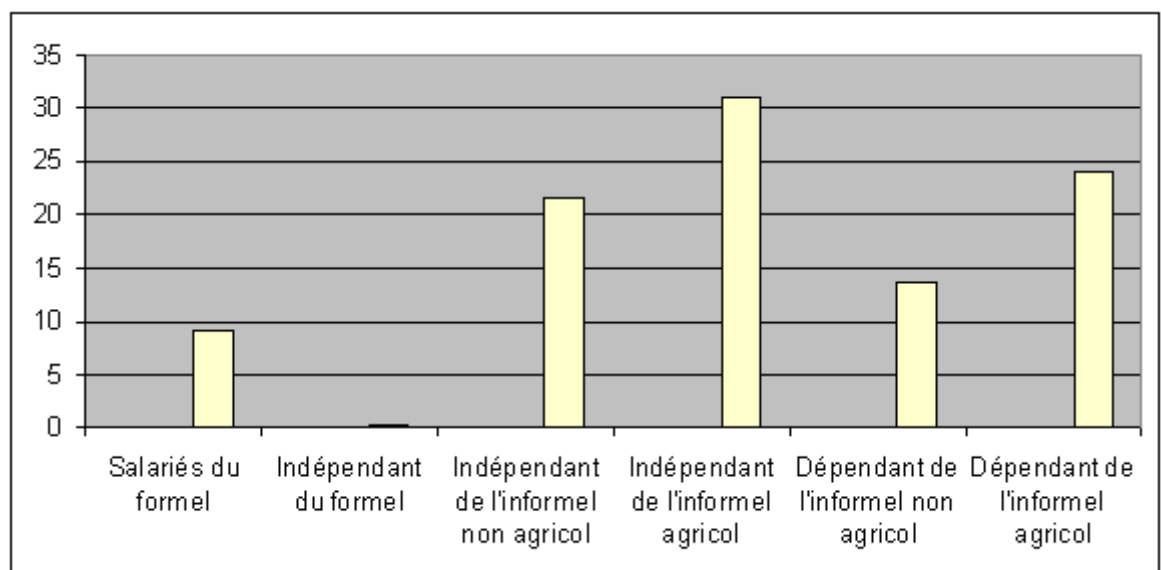


FIG. 4.6 – Emploi par statut des travailleurs

Vu le résultats présentés dans le tableau suivant, on peut dire que la proportion des femmes travaillant pour leur propre compte ou aidant leur famille est nettement supérieure à celle des hommes, ce qui confirme toujours le fait que la femme a plus tendance à jouer son rôle de mère dans le foyer. Cependant, que ce soit les postes de cadre supérieur, cadre moyen, d'employer employeur ou de manœuvre, les hommes sont plus représentés que les femmes. On note également que l'écart entre la proportion d'apprentis des deux sexes est moins important. On note également la prédominance des femmes dans l'emploi indépendant informel, ce qui pourrait indiquer qu'elles ont une forte aspiration à travailler. Les résultats révèlent également qu'elles subissent une discrimination dans l'accès à l'emploi salarial (13,7 pour cent pour les hommes contre 4,6 chez les femmes, soit environ trois fois plus élevé). Cette discrimination est d'ailleurs plus magnifeste chez les indépendants du secteur formel.

TAB. 4.6 – Répartition de la population active occupée par Catégorie socio-professionnelle (CSP)

CSP	Sexe	
	Masculin	Féminin
Profession		
Cadre supérieur	2,4	0,6
Cadre moyen	3,6	1,6
Employer/ouvrier qualifié	5,6	2,0
Employer/ouvrier semi qualifié	6,2	2,0
Manœuvre	9,4	2,0
Employeur	3,7	1,2
Travailleur pour compte propre	46,5	54,6
Aide familial	19,8	34,0
Apprenti	2,5	1,9
Type emploi		
Salarié du secteur informel	13,7	4,6
Indépendant du secteur formel	0,4	0,1
Indépendant du secteur informel non agricole	18,1	25,3
Indépendant du secteur informel agricole	31,7	30,4
Dépendant du secteur informel non agricole	16,7	10,5
Dépendant du secteur informel agricole	19,3	29,1

Source : INS, EESI 2005, Phase I

Mode de rémunération

Les Résultats de l'enquête montrent que dans l'ensemble, les hommes sont mieux rémunérés que les femmes dans ce sens que près de 38% d'hommes ont un emploi à revenu fixe contre seulement 22% des femmes actives en milieu urbain. Dans le milieu rural comme urbain les femmes ont plus des emplois moins rémunérateurs que les hommes. Cet écart est d'ailleurs plus magnifeste en milieu rural, où l'on note un écart de près de 10% contre 6,2% en milieu urbain.

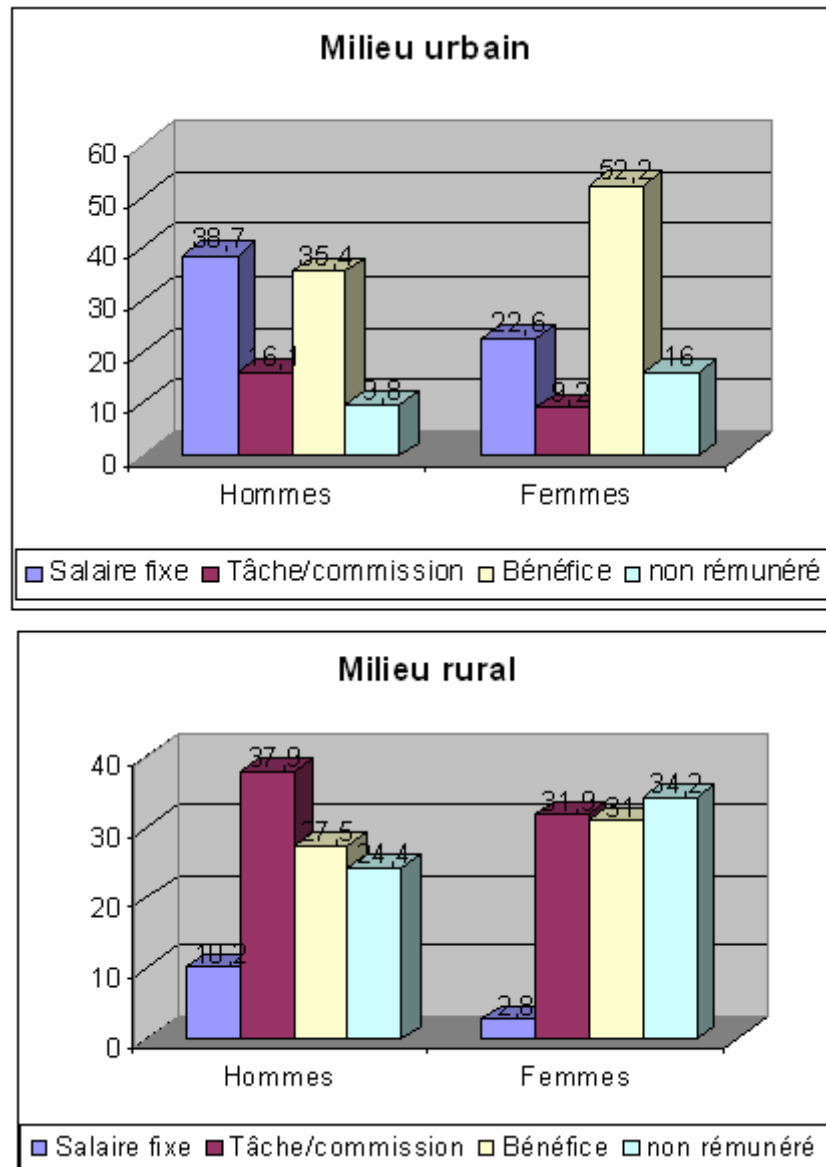


FIG. 4.7 – Répartition de la population active occupée selon le mode de rémunération

Revenu horaire

TAB. 4.7 – Répartition de la population active occupée selon le revenu horaire (en FCFA)

	Sexe	
	Masculin	Féminin
Milieu de résidence		
Urbain	403,71	241,93
Rural	169,95	84,19
Cameroun	244,89	122,48

Source : INS, EESI 2005, Phase I

Dans l'ensemble, le revenu horaire est 2 fois plus élevé chez les hommes que chez les femmes. Considérant le milieu de résidence, la même tendance est observée. Par rapport au milieu urbain, l'écart reste important, mais inférieur à celui observé en milieu rural. Par rapport à la région, on constate qu'en dehors de Douala, Yaoundé et Adamaoua ou l'écart entre le revenu horaire des hommes et des femmes est relativement moins important. Les résultats révèlent également que dans les autres régions, ce revenu est 2 fois plus élevé chez les hommes que chez les femmes. Ces résultats montrent une fois de plus qu'à temps de travail égal, les hommes sont nettement mieux rémunérés que les femmes.³

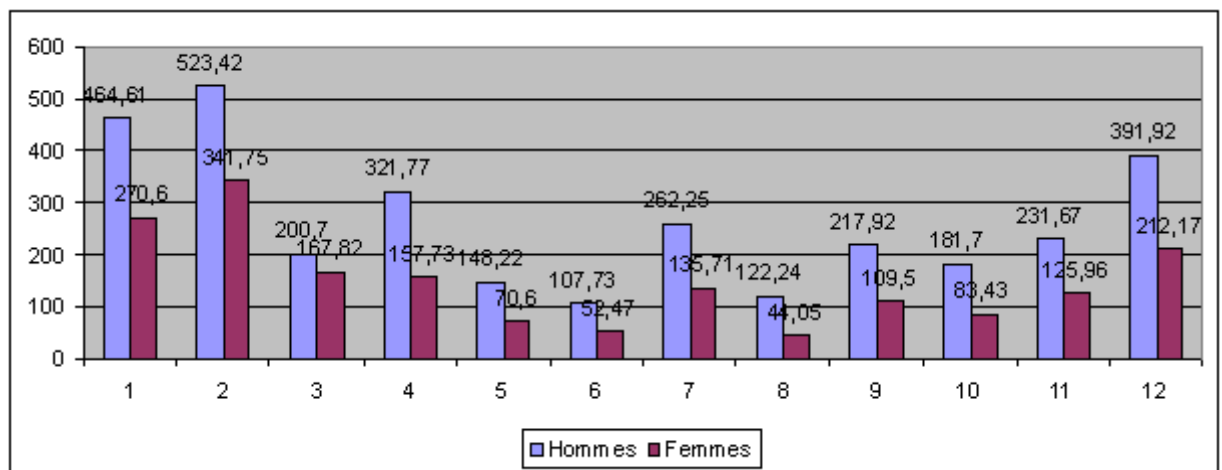


FIG. 4.8 – Répartition de la population active occupée selon leur revenu horaire (en FCFA)

Pour nous résumer, nous pouvons dire que le caractère de plus en plus informel du travail des femmes a fortement modifié leur situation sur le marché du travail. Elles exercent beaucoup plus dans des travaux à temps indépendants et à domicile. Ces travaux n'offrant cependant qu'une sécurité limitée et ne leur assurant que de faibles revenus.

³la codification de l'axe des abscisses correspond à celle utilisée à la page 11

4.2 Résultats de la mesure de la discrimination

Dans cette section, nous présentons les résultats des analyses effectuées. Tout d'abord, signalons que Le fichier de nos données étaient de type SPSS, mais puisque nous avons utilisé le logiciel Stata (version 8) pour la modélisation, ce fichier a été converti en format Stata-Dataset à l'aide du logiciel Stat-transfer.

4.2.1 Bref aperçu des tests d'hypothèses classiques utilisées lors des estimations

Tel que présenté au paragraphe 3.4.2.3, un test d'hypothèse consistera à vérifier si l'effet marginal d'une variable indépendante sur la variable dépendante est nul ou non, cela en comparant une statistique de test calculée à l'aide de paramètres estimés à une statistique⁴ critique. Cependant, le principe général d'un test est de formuler l'hypothèse que l'on veut tester (généralement appelée hypothèse nulle et notée \mathcal{H}_0) et une autre qui l'est généralement contraire (l'hypothèse alternative notée \mathcal{H}_1), cela, grâce aux observations que l'on a à sa disposition. La formulation de l'hypothèse alternative est d'une importance capitale de part de son influence sur la zone de rejet⁵ de \mathcal{H}_0 . Cette zone sera généralement déterminée en fonction du niveau de confiance⁶ choisi. Le test sera alors d'autant précis quand son niveau sera élevé. En pratique, on se fixe généralement un niveau de 5 pour cent.

Outre les statistiques de Wald et du rapport de vraisemblance utilisées lors de certains tests, nous utiliserons également la statistique t de Student (cas de l'estimation des paramètres de la fonction de gain), la statistique z de la distribution normale standard (cas des test de significativité des paramètres de la fonction de participation au marché du travail), la statistique F de Fisher (significativité globale de la fonction de gains) ou la p_value . Celle-ci est la probabilité sous $\mathcal{H}_0 : \beta = 0$ d'obtenir la valeur trouvée. **En pratique, lorsque la p_value est sous le seuil $\alpha = 5\%$ désiré, on rejètera \mathcal{H}_0 et une p_value de 0.000 permettra de le rejeter encore très fortement .**

4.2.2 Estimation des modèles

Ici, il s'agit d'estimer les paramètres des équations (3.12), (3.14) et (3.16). Notons que le choix des variables d'intérêts s'est fait non seulement suivant les critères présentés au tableau 2.2, mais également à partir d'une analyse multivariée guidée par la revue de la littérature et de nos propres connaissances. Dans les statistiques descriptives, le test de Khi-deux a été utilisé pour tester (séparément) la liaison entre un caractère et les variables dépendantes. Cela nous a permis de réduire les variables candidates afin d'éviter le risque de multi-colinéarité et de surestimation. Nous avons ainsi retenu les variables potentielles suivantes communes pour les hommes et les femmes :

⁴La statistique d'un test est une v.a.r., fonction de l'échantillon de l'étude et dont on connaît sa loi sous \mathcal{H}_0 .

⁵Elle est aussi appelée région critique et désigne par extension, l'ensemble des valeurs de décisions pour lesquelles \mathcal{H}_0 est rejetée.

⁶elle est aussi appelée niveau de test.

Cas des fonctions de participation au marché du travail

nivinst : cette variable détermine le niveau d'étude de l'individu considéré et a quatre modalités (non scolarisé, primaire, secondaire premier cycle, secondaire second cycle et supérieur) ;

age, elle détermine l'âge de l'individu ;

migrant : statut migratoire de l'individu : 0 s'il n'est pas migrant et 1 sinon ;

situation_matrimoniale : statut matrimonial, elle vaut 0 si l'individu n'est pas marié et 1 sinon) ;

aumoins_le_bac, variable dichotomique valant 1 si l'individu considéré a le Baccalauréat et 0 sinon ;

taille_menage : nombre de personnes du ménage dans laquelle fait partie l'individu ;

lien_parente : elle vaut 1 si l'individu considéré a un lien de parenté avec le chef de ménage ou son conjointe et 0 sinon ;

nombre_enfants_bas_age, c'est la variable qui détermine le nombre d'enfant de moins de 10 ans dans le ménage dont fait partie l'individu ;

travaux_domestiques : nombre d'heures de travaux domestiques que l'individu effectue par semaine.

Vu les équations (3.12), nous postulons que les modèles s'écriront :

$$\Phi^{-1}(p_i) = \beta_0 + \beta_1 \text{nivinst} + \beta_2 \text{age} + \beta_3 \text{migrant} + \beta_4 \text{situation_matrimonial} + \beta_5 \text{aumoins_bac} + \beta_6 \text{taille_menage} + \beta_7 \text{lien_parente} + \beta_8 \text{nombre_enfants_bas_age} + \beta_9 \text{travaux_domestiques} + \varepsilon_i,$$

et quant aux équations (3.16), après prédictions des revenus de réserve données par les équations (3.14), chaque modèle s'écrira :

$$\Phi^{-1}(p_i) = \beta_0 + \beta_1 \text{nivinst} + \beta_2 \text{age} + \beta_3 \text{migrant} + \beta_4 \text{situation_matrimonial} + \beta_5 \text{aumoins_bac} + \beta_6 \text{taille_menage} + \beta_7 \text{lien_parente} + \beta_8 \text{nombre_enfants_bas_age} + \beta_9 \text{travaux_domestiques} + \beta_{10} \text{Revenu} + \varepsilon_i,$$

Cas des fonctions de gains

Outre les variables **nivinst** et **migrant** précédentes, nous avons également les variables suivantes :

t3_5_travailleurs, elle vaut 1 si le nombre de travailleurs de l'entreprise dans laquelle travaille l'enquêté est compris entre 3 et 5, et 0 sinon ;

plus_5_travailleurs, elle vaut 1 si l'entreprise dans laquelle travaille l'enquêté a plus de 5 travailleurs et 0 sinon ;

experience_potentielle variable **ap8a1** de la base, détermine le nombre d'années que l'enquêté a déjà mit dans l'entreprise ;

experience_emploi, c'est la variable **ap8a2** de la base, elle détermine le nombre d'années que l'individu a déjà mit dans l'emploi qu'il exerce ;

nationalité, cette variable est obtenue par codification de la variable **m7** de la base, libellé nationalité qui elle, avait 9 modalités : Cameroun, Nigeria, autre pays de la cemas, autre pays de l’afrique de l’ouest, autre pays d’afrique, europe, chine, autre d’asie et amérique, reste du monde ;

paie_impôt, Cette variable permet de vérifier si l’entreprise dans laquelle travaille l’enquêté est de type formelle, i.e. possède un numéro de contribuable ; elle est créée à partir de la variable bulletin_paie de la base.

Le modèle s’écrit ainsi après évaluation de l’inverse de ratio de Mills (*inv_mills_hf*) :

$$LnRevenu = \beta_0 + \beta_1nivist + \beta_2 experience_emploi + \beta_3 experience_potentielle + \beta_4migrant + \beta_5nationalite + \beta_6t3_5_travailleurs + \beta_7plus_5_travailleurs + \beta_8paie_impot + \beta_9inv_mills_hf + \varepsilon_i$$

1-) Estimation des paramètres des équations (3.12)

***) cas des hommes**

TAB. 4.8 – Estimation probit de la fonction de participation des hommes au marché du travail : modèle permettant d’estimer le vecteur ratio de mills conditionné par les hommes

variable dépendante	vardpte = actif occupée				
variables explicatives	coefficients	erreur standard	p-value	I.C. à 95 pour cent	
constante	1.647	0.170	0.000	1.315	1.980
non scolarisé	réf	réf	réf	réf	réf
primaire	-0.064	0.155	0.677	-0.368	0.239
secondaire 1er cycle	-0.348	0.155	0.025	-0.652	-0.044
secondaire 2nd cycle	-0.571	0.142	0.000	-0.849	0.293
supérieur	-0.897	0.111	0.000	-1.116	-0.679
Bac ou plus	0.244	0.120	0.041	0.010	0.479
âge	0.011	0.022	0.000	-1.015	-.0064
migrant	-0.116	0.045	0.011	-0.204	-0.026
situation matrimoniale	0.282	0.066	0.000	0.153	0.411
taille ménage	0.033	0.128	0.009	0.008	0.059
lien de parenté	0.761	0.066	0.000	0.632	0.891
nombre d’enfants en bas age	-0.046	0.016	0.005	-0.075	-0.014
travaux domestiques	-0.016	0.003	0.000	1.315	-1.011

Nombre d’observations = 9193

LR chi2(12) = 511.42 (Prob > chi2 = 0.0000)

Hosmer-Lemeshow chi2(1997) = 1959.34 (Prob > chi2 = 0.7221)

Interprétation statistique du tableau 4.8 :

La colonne coefficients de ce tableau correspond aux paramètres des équations (3.12), ce sont donc les coefficients de régression dans le cas du sexe masculin, nous pouvons alors écrire le modèle comme suit :

$$\Phi^{-1}(p_i) = 1.647 - 0.064(\text{nivinst} = \text{primaire}) - 0.348(\text{nivinst} = \text{secondaire1ercycle}) - 0.571(\text{nivinst} = \text{secondaire2ndcycle}) - 0.897(\text{nivinst} = \text{suprieur}) + 0.244\text{aumoins_bac} + 0.011\text{age} - 0.116\text{migrant} + 0.282\text{situation_matrimonial} + 0.033\text{taille_menage} + 0.761\text{lien_parente} - 0.046\text{nombre_enfants_bas_age} - 0.016\text{travaux_domestiques} + \varepsilon_i,$$

où $p_i = Pr(\text{vardpte} = 1)$,

et par ailleurs, ce modèle est obtenu sur un échantillon de 9193 individus (Nombre d'observations = 9193). LR $\chi^2(12) = 511.42$ signifie que la valeur de la statistique du test de rapport de vraisemblance (LR) est 511.42 et à un seuil de 5 pour cent, elle est nettement très significatif ($Prob > \chi^2 = 0.0000$). Ce qui permet ainsi de dire que les variables explicatives intervenant dans ce modèle prises ensemble ont un effet très positif sur la variable dépendante. **En d'autres termes, on dit que le modèle est globalement significatif.**

La colonne erreur standard correspond à l'écart type de chaque variable indépendante. Cette erreur mesure la précision d'un coefficient ; plus elle est petite, plus la précision est bonne et a fortiori le coefficient est plus efficace. Le test de significativité individuelle est basé sur cette erreur. En effet, la statistique utilisée ici, est la z de la distribution normale standard (évoquée en début de section) et est par définition égale à $\frac{\text{coefficient}}{\text{ecart type}}$.

L'avant dernière colonne de ce tableau correspond à la p_value associée à z . Sous Stata, elle est programmé de sorte que lorsqu'on a $p_value > 0.05$ (0.05 étant le seuil choisi) alors, la variable indépendante correspondant est statistiquement significatif, sinon il n'y a pas d'évidence que cette variable ait un effet sur la variable dépendante.

La dernière colonne donne l'intervalle de confiance à 95% du coefficient de la ligne correspondante.

a) Calibration du modèle : Test de Hosmer et Lemeshow

Ce test tel que décrit au paragraphe précédent est basé sur un regroupement des probabilités prédites par le modèle (par exemple par décile, quintile, percentile). On calcule ensuite pour chaque groupe le nombre observé de réponses positives (**vardpte = 1**) et négative (**vardpte = 0**) que l'on compare au nombre espéré prédit par le modèle. Par suite, on calcule alors une distance entre les fréquences observées et prédites au moyen d'une statistique de Khi-2 et lorsque la distance est petite, on considère que le modèle est bien calibré.

Ce test est passé dans le cas de ce modèle ($\chi^2(1997) = 1959.34$ ($Prob > \chi^2 = 0.7221$)) pour un regroupement de 1999 quintiles, ce qui nous permet en d'autres termes, de dire que l'ajustement global du modèle à notre jeu de données est satisfaisant.

b) Adéquation du modèle

Pour ce modèle, l'aire sous la courbe ROC vaut 74,25%, ce qui traduit un pouvoir discriminant acceptable.

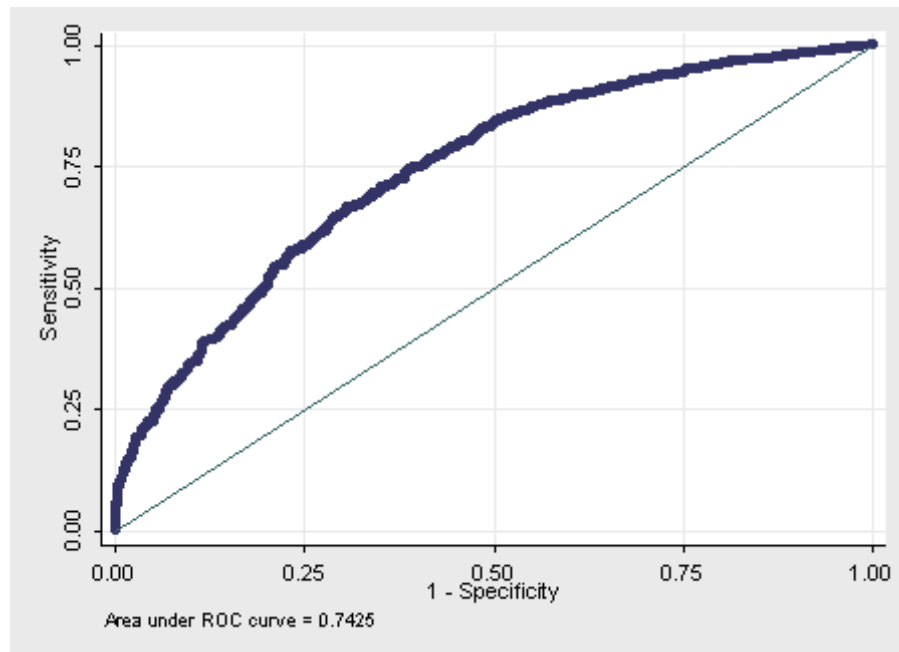


FIG. 4.9 – Pouvoir de discrimination : aire sous la courbe ROC (Hommes)

Le seuil de discrimination est de 0,92 (point où la sensibilité est (ou presque) égale à la spécificité, il est obtenu à l'aide de la commande stata suivante `lstat, cutoff(0.92)`), contrairement à la valeur 0,5 généralement prise par défaut. A ce seuil, la Sensibilité vaut 68,12% et la spécificité est de 66.81 %.

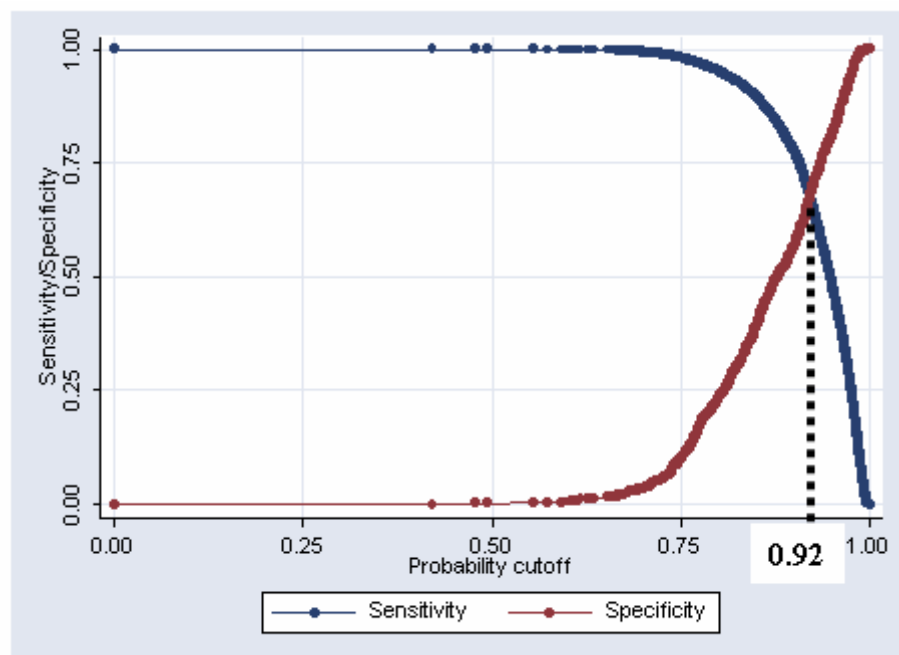


FIG. 4.10 – Courbe de la sensibilité et de la spécificité en fonction des probabilités prédites (Hommes)

c) Validation des probabilités prédites

La validation des probabilités prédites a été obtenue grâce à la fonction `linktest` de `stata`. Cependant, elle ne teste que l'hypothèse selon laquelle les probabilités prédites sont sans biais sans toute fois préciser le taux de bonne prédiction (ou le taux d'erreur), ce qui nous a amené à construire un taux de prédiction à partir de la césure 0,92 (donnée par l'intersection de la courbe de la sensibilité et de la spécificité précédente). **Notons que cette démarche est légitime, car la variable dépendante est dichotomique.**

Démarche utilisée :

- 1- Estimation du modèle.
- 2- Prédiction des probabilités \hat{p} pour tous les individus.
- 3- Création d'une variable dichotomique \hat{y} valent 1 si $\hat{p} \geq \text{cesure}$ et 0, sinon.
- 4- Création d'une variable de comparaison : $T = \text{vardpte} - \hat{y}$

Les valeurs possibles de T sont alors -1 (cas où $\text{vardpte}=0$ et $\hat{y} = 1$, correspondant à un cas de mauvaise prédiction), 0 ($\text{vardpte} = \hat{y}$, **cas de bonne prédiction**) ou 1 (cas où $\text{vardpte}=1$ et $\hat{y} = 0$, qui correspond également à un cas de mauvaise prédiction).

- 5- *Taux de bonne prédiction* = $\frac{T}{N}$, où N est la taille de l'échantillon effectivement utilisée dans l'estimation.

Nous avons ainsi obtenu un taux de bonne prédiction de l'ordre de 88,52%, ce qui est du point de vue statistique acceptable.

d) Analyse des résidus

En plus des quatre études précédemment effectuées, une analyse des résidues est réalisée afin de vérifier qu'il n'y a pas d'erreur systématique ou des observations aberrantes (i.e. très mal expliquées). Mais, le graphe des résidus ci-dessus construit ne nous présente pas un très grand déséquilibre. **Au regard de toutes ces analyses, nous pouvons donc conclure que la qualité du modèle retenu est globalement satisfaisant.**

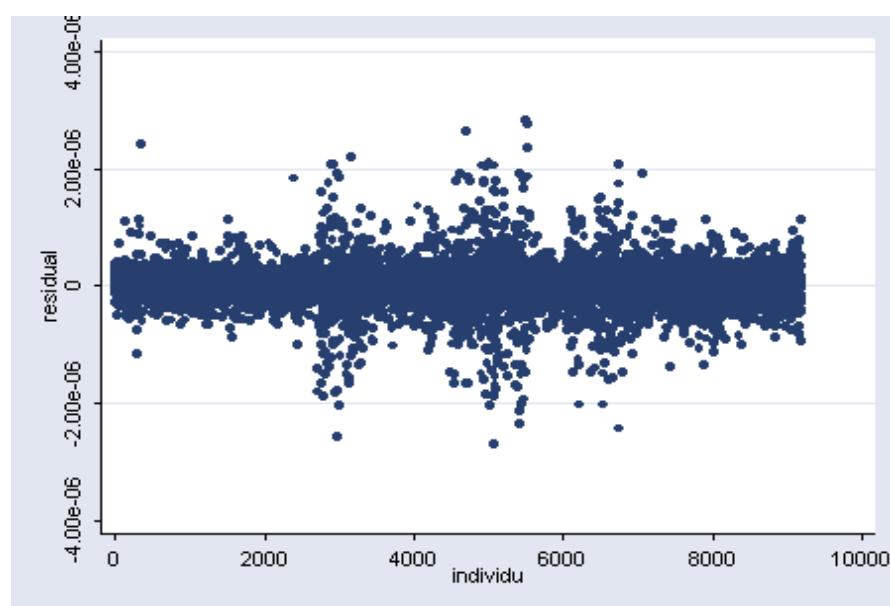


FIG. 4.11 – Graphe des résidus

****) cas des femmes**

D’après les remarques faites précédentes le tableau ci-dessus nous permet d’écrire l’équation (3.12) selon le sexe féminin comme suit :

$$\Phi^{-1}(p_i) = 0.983 - 0.154(\text{nivinst} = \text{primaire}) - 0.602(\text{nivinst} = \text{secondaire1}^{\text{er}} \text{cycle}) - 0.912(\text{nivinst} = \text{secondaire2}^{\text{nd}} \text{cycle}) - 1.179(\text{nivinst} = \text{suprieur}) + 0.196\text{aumoins_bac} + 0.023\text{age} - 0.239\text{migrant} + 0.037\text{situation_matrimonial} + 0.064\text{taille_menage} + 0.048\text{lien_parente} - 0.077\text{nombre_enfants_bas_age} - 0.017\text{travaux_domestiques} + \varepsilon_i,$$

où $p_i = Pr(\text{vardpte} = 1)$,

TAB. 4.9 – Estimation probit de la fonction de participation des femmes au marché du travail : modèle permettant d’estimer le vecteur ratio de mills conditionné par les femmes

variable dépendante	vardpte = actif occupée				
variables explicative	coefficients	erreur standard	p-value	I.C. à 95 pour cent	
constante	0.983	0.160	0.000	0.670	1.297
non scolarisé	réf	réf	réf	réf	réf
primaire	-0.154	0.142	0.281	-0.432	.125
secondaire 1er cycle	-0.602	0.142	0.000	-0.881	-0.322
secondaire 2nd cycle	-0.912	0.129	0.000	-1.165	-0.660
supérieur	-1.179	0.1072	0.000	-1.389	-0.969
Bac ou plus	0.196	0.120	0.102	-0.039	0.430
âge	0.023	0.002	0.000	0.021	0.030
migrant	-0.239	0.043	0.000	-0.323	-0.155
situation matrimoniale	0.037	0.051	0.470	-0.063	0.137
taille ménage	0.064	0.014	0.000	0.038	0.091
lien de parenté	0.048	0.062	0.433	-0.072	0.169
nombre d’enfant en bas age	-0.077	0.018	0.000	-0.112	-0.042
travaux domestiques	-0.017	0.001	0.000	-0.014	-0.009

Nombre d’observations = 8142

LR chi2(12) = 757.28

Prob > chi2 = 0.0000

a) Calibration du modèle : Test de Hosmer et Lemeshow

Ce test est également satisfaisant pour ce modèle ($\text{chi2}(1998) = 1855, 83$ ($\text{Prob} > \text{chi2} = 0, 9891$)) pour un regroupement de 2000 quintiles, ce qui nous permet aussi de dire que l’ajustement global du modèle est satisfaisant.

b) Adéquation du modèle

L’aire sous la courbe ROC vaut 75,77%, ce qui est également équivalent à un pouvoir discriminant acceptable.

Le seuil de discrimination est de 0,86. A ce seuil, la sensibilité vaut 69,06% et la spécificité est de 67,81 %.

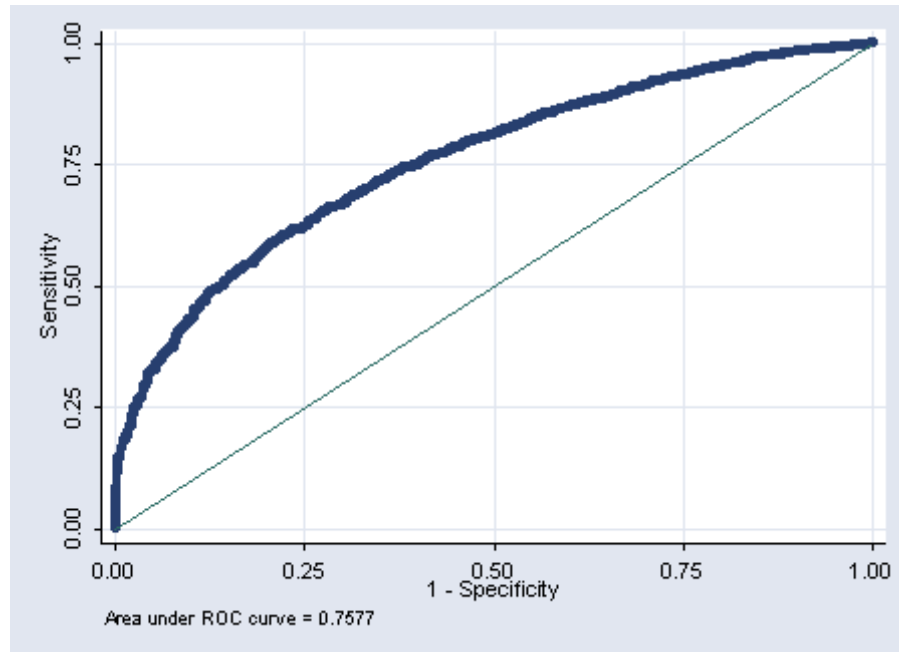


FIG. 4.12 – Pouvoir de discrimination : aire sous la courbe ROC (Femmes)

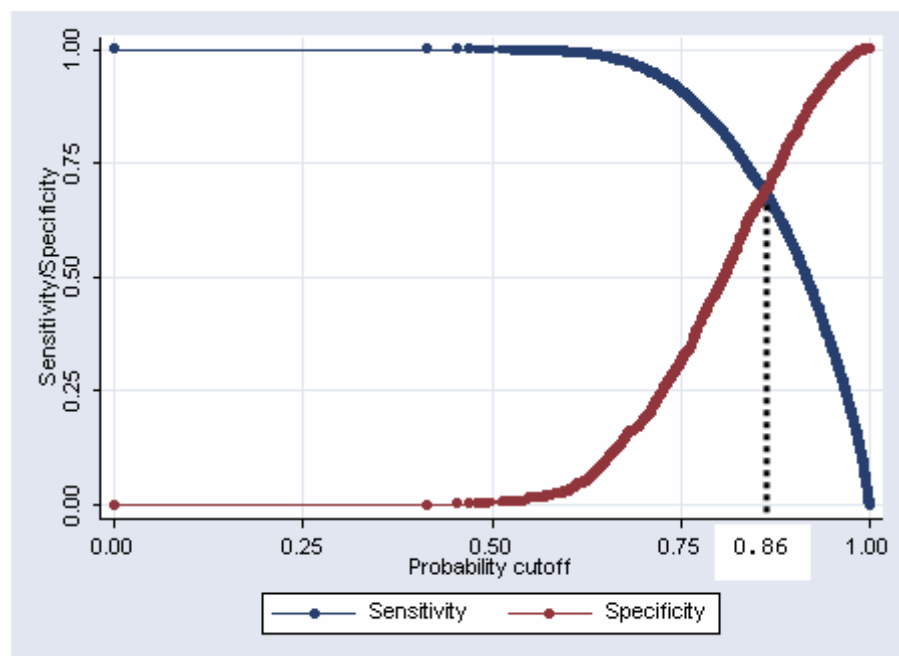


FIG. 4.13 – Courbe de la sensibilité et de la spécificité en fonction des probabilités prédites (Femmes)

c) Validation des probabilités prédites

Nous avons obtenu un taux de bonne prédiction de l'ordre de 78,04%, ce qui est du point de vue statistique également acceptable.

d) Analyse des résidus

L'analyse des résidus ne nous présente également pas un très grand déséquilibre. La plupart des résidus sont concentrés autour de la droite d'équation $y = 0$. **Nous pouvons alors conclure quant à cette étude et vues toutes les analyses précédentes, que la qualité du modèle retenu est globalement satisfaisant.**

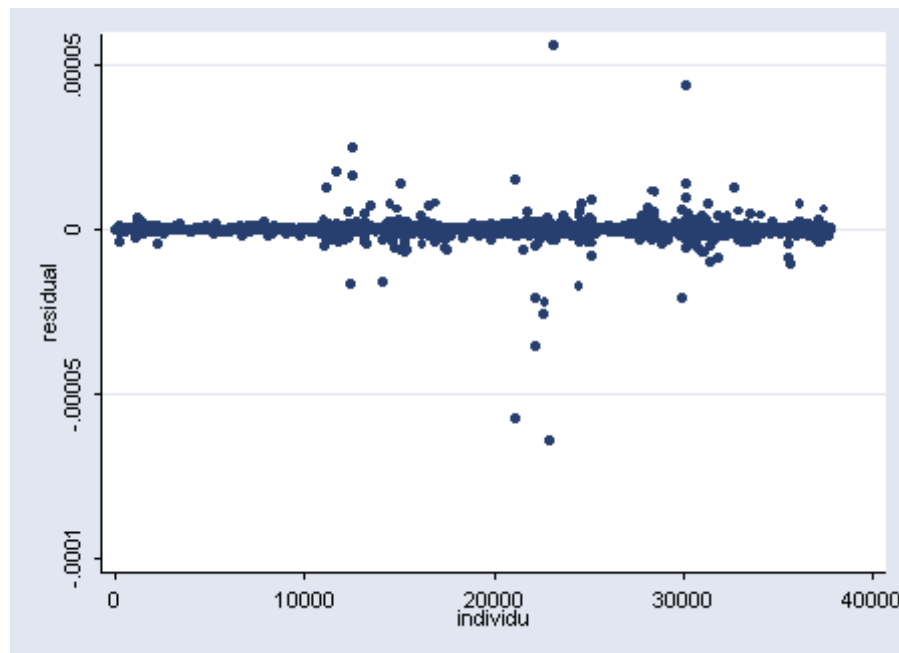


FIG. 4.14 – Graphe des résidus

Nous pouvons à présent estimer les coefficients du vecteur ratio de Mills qui nous permet de corriger le biais de sélection lié à l'estimation des équations de gains.

2-) Estimation des paramètres des équations (3.14) : Fonctions de gains

*) cas des hommes

Le modèle suivant a été retenu pour la fonction de gain des hommes (tableau 4.10) :

$$\begin{aligned} \text{Ln}(\text{Revenu}) = & 2.768 + 0.256(\text{nivinst} = \text{primaire}) + 0.645(\text{nivinst} = \text{secondaire1ercycle}) + \\ & + 1.015(\text{nivinst} = \text{secondaire2ndcycle}) + 1.712(\text{nivinst} = \text{suprieur}) - 0.040\text{experience_emploi} + \\ & + 0.045\text{experience_potentielle} + 0.314\text{migrant} - 0.161\text{nationalité} + 0.036\text{t_3_5travailleurs} + \\ & + 0.598\text{plusdetravailleurs} + 0.387\text{paie_impot} - 2.130\text{ratio_de_Mills} + \varepsilon_i \end{aligned}$$

Le test F de Fisher montre que les variables explicatives choisies, prises en ensemble sont significativement différentes de zéro au seuil de 5% et globalement, le modèle retenu explique 45,08% de la variance du logarithme des gains.

Ce modèle a été retenu grâce à l'analyse des résidus. En effet, la suppression des valeurs aberrantes a permis d'obtenir près de 95% de valeurs résiduelles (voir figure 4.15) dans l'intervalle $[-2; 2]$ et a eu un effet positif sur la valeur de la statistique R^2 qui est passée de 38,43% à 45,08%. La suppression des observations ne pourrait en aucun cas mettre en cause la qualité de notre ajustement, puisque nos données sont issues d'une enquête.

TAB. 4.10 – Estimation des coefficients de la fonction de gains des hommes

variable dépendante	$Ln(\text{Revenu})$				
variables explicative	coefficients	erreur standard	statistique (t)	I.C. à 95 pour cent	
constante	2.768	0.063	(***)	2.643	2.894
non scolarisé	réf	réf	réf	réf	réf
primaire	0.256	0.032	(***)	0.194	0.319
secondaire 1er cycle	0.645	0.036	(***)	0.575	0.715
secondaire 2nd cycle	1.015	0.040	(***)	0.936	1.093
supérieur	1.712	0.048	(***)	1.618	1.805
expérience_emploi	-0.040	0.004	(***)	-0.047	-0.034
expérience_potentielle	0.045	0.003	(***)	0.039	0.052
migrant	0.314	0.020	(***)	0.274	0.353
nationalité	-0.161	0.058	(**)	-0.274	-0.048
t_3_5 travailleurs	0.036	0.026	(-)	-0.016	0.087
plus de 5 travailleurs	0.598	0.023	(***)	0.553	0.643
paie_impôt	0.387	0.021	(***)	0.347	0.430
ratio_de_Mills	-2.130	0.110	(***)	-2.346	-1.915

Nombre d'observations = 7063 F(12, 7296) = 482.20(***)
R-squared = 0.4508
(***) = $p_value < 0.001$; (**) = $p_value < 0.01$; (*) = $p_value < 0.05$
(-) = non significatif

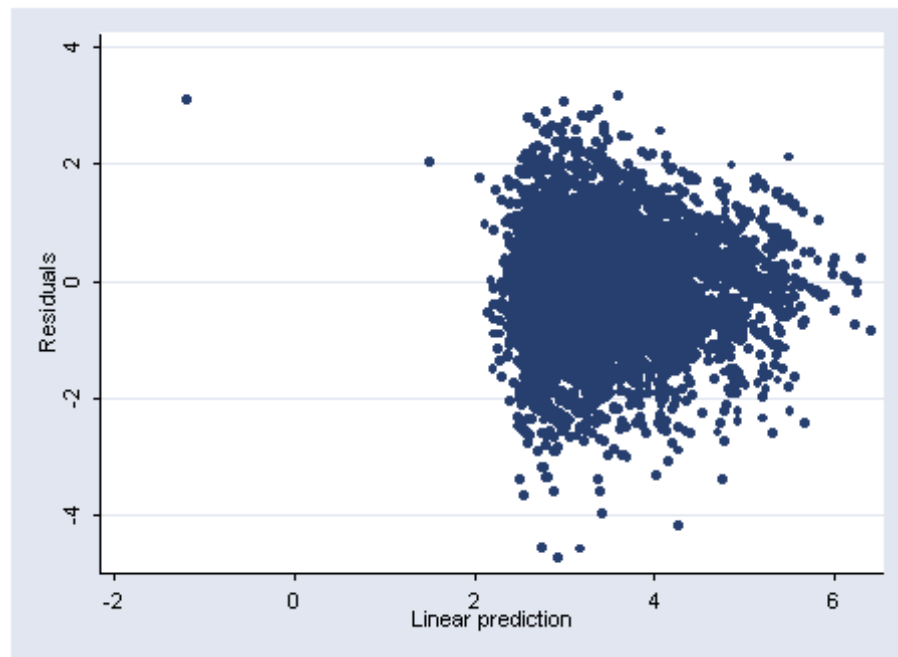


FIG. 4.15 – Graphe des résidus (Hommes)

****) cas des femmes**

Le modèle suivant a été retenu pour la fonction de gain des hommes (tableau 4.10) :

$$\begin{aligned} \text{Ln}(\text{Revenu}) = & 2.187 + 0.570(\text{nivinst} = \text{primaire}) + 1.177(\text{nivinst} = \text{secondaire1ercycle}) + \\ & + 1.848(\text{nivinst} = \text{secondaire2ndcycle}) + 2.671(\text{nivinst} = \text{suprieur}) - 0.037\text{experience_emploi} + \\ & + 0.039\text{experience_potentielle} + 0.293\text{migrant} - 0.143\text{nationalité} - 0.022\text{t_3_5travailleurs} + \\ & + 0.639\text{plusde travailleurs} + 0.514\text{paie_impot} - 2.222\text{ratio_de_Mills} + \varepsilon_i \end{aligned}$$

Pour ce modèle, le test *F* de Fisher montre que les variables explicatives choisies, prises en ensemble sont significativement différentes de zéro au seuil de 5% et globalement, le modèle retenu explique 46, 53% de la variance du logarithme des gains chez les femmes.

L'analyse des résidus à travers la suppression des valeurs aberrantes a permis d'obtenir près de 95% de valeurs résiduelles (**voir figure 4.16**) dans l'intervalle $[-2; 2]$ et a eu un effet positif sur la valeur de la statistique R^2 qui est passé de 39, 66% à 45, 08%.

TAB. 4.11 – Estimation des coefficients de la fonction de gains des femmes

variable dépendante	Ln(Revenu)				
variables explicatives	coefficients	erreur standard	statistique (t)	I.C. à 95 pour cent	
constante	2.187	0.088	(***)	2.014	2.360
non scolarisé	réf	réf	réf	réf	réf
primaire	0.570	0.030	(***)	0.512	0.629
secondaire 1er cycle	1.177	0.043	(***)	1.093	1.262
secondaire 2nd cycle	1.848	0.060	(***)	1.731	1.965
supérieur	2.671	0.078	(***)	2.519	2.823
expérience_emploi	-0.037	0.005	(***)	-0.048	-0.027
expérience_potentielle	0.039	0.005	(***)	0.028	0.049
migrant	0.293	0.023	(***)	0.248	0.337
nationalité	-0.143	0.085	(-)	-0.310	0.023
t_3_5 travailleurs	-0.022	0.029	(-)	-0.080	0.036
plus de 5 travailleurs	0.639	0.034	(***)	0.572	0.705
paie_impôt	0.514	0.033	(***)	0.449	0.578
ratio_de_Mills	-2.222	0.126	(***)	-2.470	-1.974

Nombre d'observations = 5369

F(12, 7296) = 388.46 (***)

R-squared = 0.4653

(***) = $p_value < 0.001$; (**) = $p_value < 0.01$; (*) = $p_value < 0.05$

(-)=non significatif

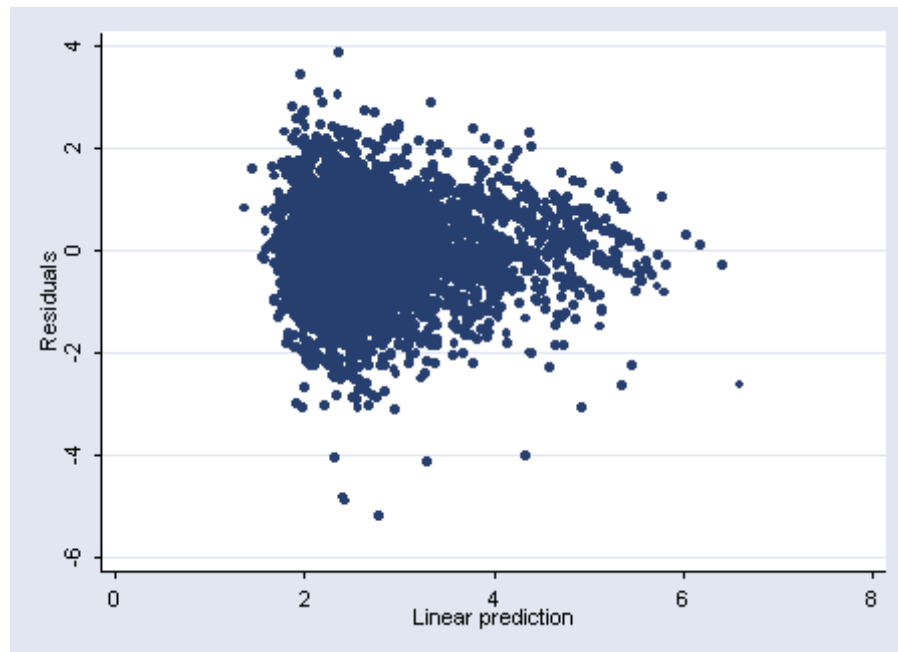


FIG. 4.16 – Graphe des résidus (Femmes)

***) Analyse comparée des résultats des estimations : fonctions de gains

Le tableau 4.12 ci-dessous donne un récapitulatif des estimations des fonctions de gains. Le test de l'égalité des coefficients qui prend la valeur 66.57 nous a permis de rejeter l'hypothèse nulle selon laquelle une seule équation expliquerait les gains des hommes et des femmes. Comme signalé un peu plus haut, les tests t ont pour chaque sexe, montré que les coefficients pris en groupe sont significativement différents de zéro au seuil de 5%. Les deux modèles montrent que le coefficient de l'inverse du ratio de Mills est négativement significatif pour les hommes comme pour les femmes. Cela justifie l'existence d'un biais d'auto-sélection associé à l'échantillon utilisé. En d'autres termes, on dit que le processus de participation des individus au marché du travail n'est pas aléatoire.

L'analyse de ce résultat nous montre également que, le nombre d'années qu'un individu met dans une entreprise augmente systématiquement son gain, contrairement au temps qu'il a mis dans son emploi, qui plutôt, diminue substantiellement son gain. On note également que le fait qu'un individu occupe un emploi dans une entreprise de moins de cinq employés n'influence pas son gain contrairement au fait d'être dans une entreprise employant plus de cinq travailleurs. En d'autres termes, le gain augmente en fonction de la taille de l'entreprise. De même, pour les hommes comme pour les femmes, Le fait d'occuper un emploi dans une entreprise du secteur formel a un effet positif sur son revenu. En outre, un migrant gagne 0,314 environ, plus qu'un non migrant ; chez les femmes, l'augmentation est d'environ 29,3%. Enfin, on note que les variables d'éducation ont une influence positive sur les gains des individus par rapport à la variable de référence, sans éducation. La fonction de gain est croissante avec le niveau d'instruction.

En définitive, on a deux modèles globalement significatives qui expliquent respectivement 45,08% et 46,05% des gains des hommes et des femmes. Elles pourront donc être utilisées pour évaluer le degré de la discrimination salariale sur le marché du travail.

TAB. 4.12 – Récapitulatif des résultats des estimations : fonctions de gains

variable dépendante	$Ln(\text{Revenu})$			
	Hommes		Femmes	
variables explicatives	coefficients	Statistique (t)	coefficients	Statistique(t)
constante	2.768	(***)	2.187	(***)
non scolarisé	réf	réf	réf	réf
primaire	0.256	(***)	0.570	(***)
secondaire 1er cycle	0.645	(***)	1.177	(***)
secondaire 2nd cycle	1.015	(***)	1.848	(***)
supérieur	1.712	(***)	2.671	(***)
expérience_emploi	-0.040	(***)	-0.037	(***)
expérience_potentielle	0.045	(***)	0.039	(**)
migrant	0.314	(***)	0.293	(***)
nationalité	-0.161	(**)	-0.143	(-)
t_3_5 travailleurs	0.036	(-)	-0.022	(-)
plus de 5 travailleurs	0.598	(***)	0.639	(***)
paie_impôt	0.387	(***)	0.514	(***)
ratio_de_Mills	-2.130	(***)	-2.222	(***)
Nombre d'observations	7063		5369	
R-squared	0.4508		0.4653	
Statistique de Fisher	482.20 (***)		388.46 (***)	
Test de l'égalité des coefficients : valeur de la statistique = 66.57 (***)				
(***) = $p_value < 0.001$; (**) = $p_value < 0.01$; (*) = $p_value < 0.05$				
(-)=non significatif				

3-) Modèles finales de participation au marché du travail : Estimation des paramètres des équations (3.16)

***) cas des hommes**

TAB. 4.13 – Estimation des coefficients de la fonction de participation des hommes au marché du travail

variable dépendante	vardpte = actif occupée				
variables explicative	coefficients	erreur standard	statistique	I.C. à 95 pour cent	
constante	4.138	0.236	(***)	3.676	4.600
non scolarisé	réf	réf	réf	réf	réf
primaire	0.451	0.166	(**)	0.126	0.776
secondaire 1er cycle	0.458	0.170	(**)	0.125	0.790
secondaire 2nd cycle	0.519	0.163	(**)	0.20	0.839
supérieur	0.750	0.154	(***)	0.449	1.052
Bac ou plus	0.534	0.123	(***)	0.293	0.775
âge	-0.016	0.002	(***)	-0.020	-0.011
migrant	0.202	0.051	(***)	0.102	0.303
situation matrimoniale	0.475	0.071	(***)	0.337	0.614
taille ménage	0.056	0.013	(***)	0.030	0.082
lien de parenté	1.317	0.078	(***)	1.163	1.470
nombre d'enfant en bas age	-0.063	0.017	(***)	-0.097	-0.030
travaux domestiques	-0.029	0.003	(***)	-0.034	-0.023
Revenu de réserve	-1.052	0.064	(***)	-1.178	-0.926

Nombre d'observations = 9193

LR chi2(12) = 787.24 (***)

(***) = $p_value < 0.001$ (**) = $p_value < 0.01$; (*) = $p_value < 0.05$

a) Calibration du modèle : Test de Hosmer et Lemeshow

Ce test est passé pour ce modèle pour un regroupement de 2000 quintiles ($chi2(1998) = 1988.70$ ($Prob > chi2 = 0.5544$)), ce qui permet de valider la qualité de l'ajustement global du modèle.

b) Adéquation du modèle

L'aire sous la courbe ROC vaut 80,64%, donc le pouvoir de discrimination est **excellent**. Le seuil de discrimination est de 0,913; la sensibilité vaut 74,31% et la spécificité est de 74,01%.

c) Validation des probabilités prédites

Nous avons obtenu un taux de bonne prédiction de l'ordre de 88,75%, ce qui est excellent du point de vue statistique.

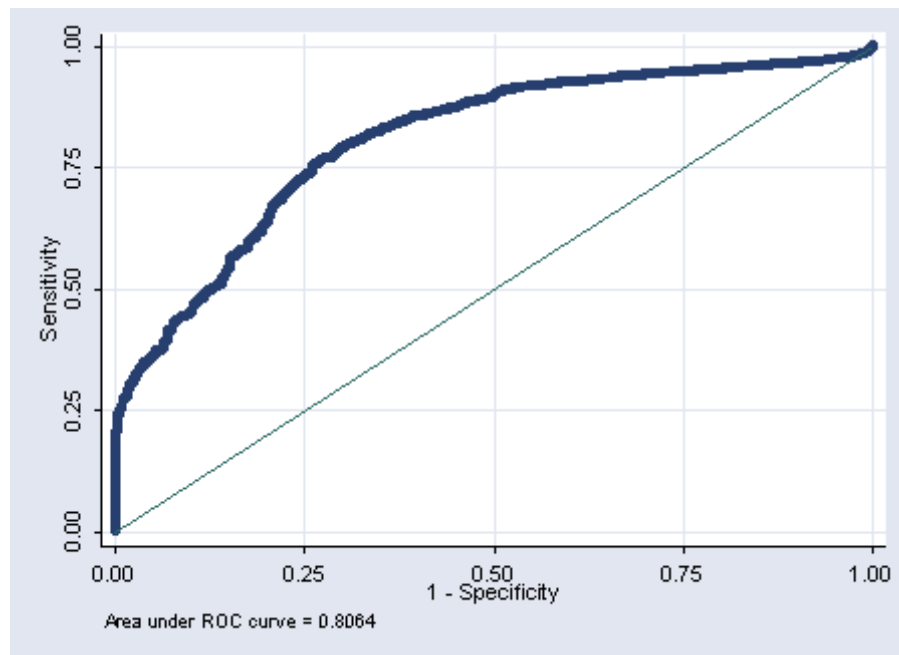


FIG. 4.17 – Pouvoir de discrimination : aire sous la courbe ROC (Hommes)

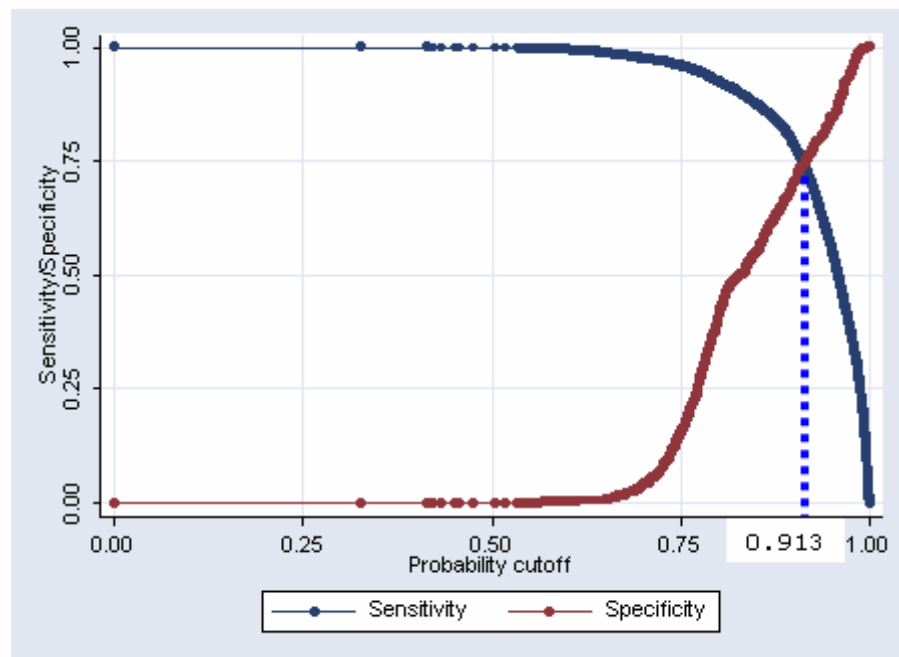


FIG. 4.18 – Courbe de la sensibilité et de la spécificité en fonction des probabilités prédites (Hommes)

**) cas des femmes

TAB. 4.14 – Estimation des coefficients de la fonction de participation des femmes au marché du travail

variable dépendante	vardpte = actif occupée				
variables explicatives	coefficients	erreur standard	statistique	I.C. à 95 pour cent	
constante	3.764	0.236	(***)	3.370	4.159
non scolarisé	réf	réf	réf	réf	réf
primaire	1.392	0.165	(***)	1.070	1.715
secondaire 1er cycle	1.652	0.172	(***)	1.314	1.990
secondaire 2nd cycle	2.335	0.175	(***)	1.992	2.679
supérieur	3.404	0.189	(***)	3.032	3.775
Bac ou plus	0.831	0.126	(***)	0.584	1.078
âge	0.070	0.003	(***)	0.064	0.077
migrant	0.082	0.050		-0.016	0.180
situation matrimoniale	0.053	0.058		-0.062	0.168
taille ménage	0.154	0.016	(**)	0.123	0.186
lien de parenté	0.134	0.071		-0.004	0.273
nombre d'enfants en bas age	-0.173	0.021	(*)	-0.213	-0.132
travaux domestiques	-0.036	0.002	(**)	-0.040	-0.032
Revenu de réserve	-2.050	0.068	(***)	-2.183	-1.916

Nombre d'observations = 8142

LR chi2(12) = 1850.67 (**)

(***) = $p_value < 0.001$ (**) = $p_value < 0.01$; (*) = $p_value < 0.05$

a) Calibration du modèle : Test de Hosmer et Lemeshow

Ce test est également passé pour ce modèle ($chi2(1998) = 2464,76$ ($Prob > chi2 = 0,6783$)), ce qui nous permet aussi de dire que l'ajustement global du modèle est satisfaisant.

b) Adéquation du modèle

L'aire sous la courbe ROC vaut 90,07%, ce qui correspond à un **excellent** pouvoir de discrimination.

Le seuil de discrimination est de 0,80 ; à ce seuil, la sensibilité vaut 84,45% et la spécificité est de 83,54%.

c) Validation des probabilités prédites

Le taux de bonne prédiction est de l'ordre de 80,18%, ce qui est excellent du point de vue statistique.

Notons enfin que pour la validation de chacun de ces deux modèles, une analyse des résidus a été effectuée et les résultats obtenus étaient similaires à ceux obtenus lors de l'estimation des équations réduites de participation qui nous ont permis d'obtenir l'inverse du ratio de Mills.

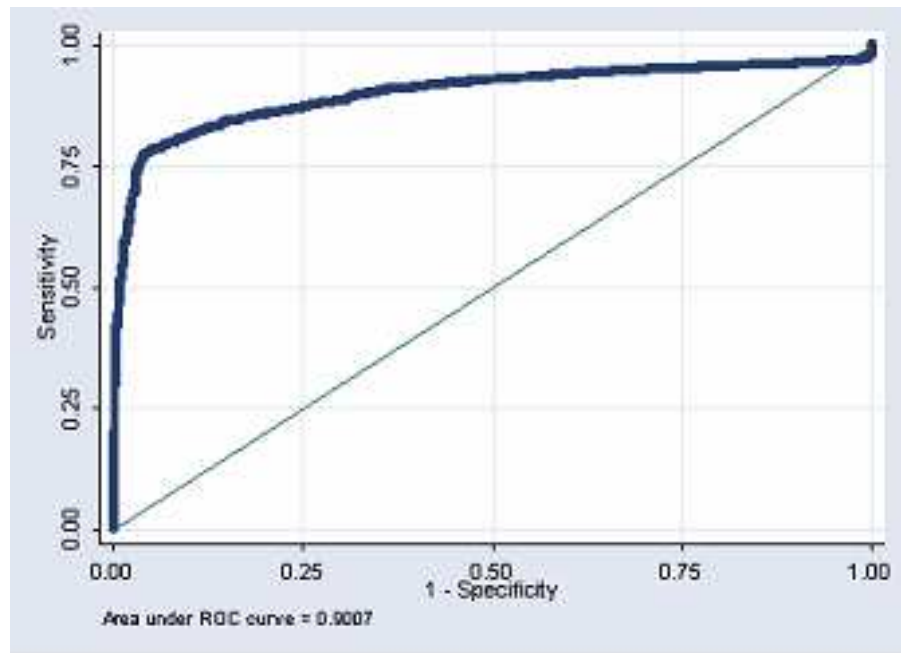


FIG. 4.19 – Pouvoir de discrimination : aire sous la courbe ROC (Femmes)

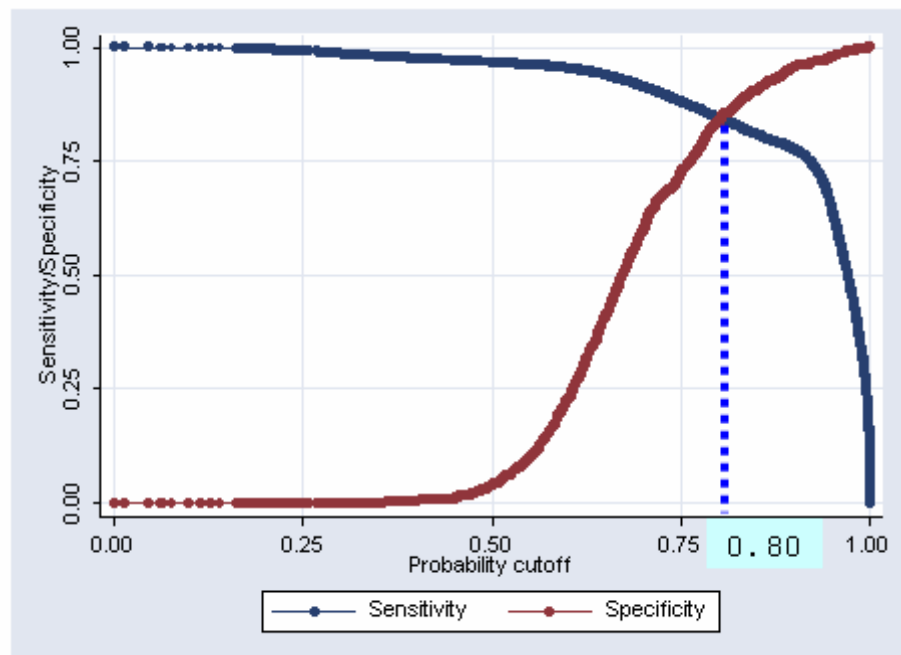


FIG. 4.20 – Courbe de la sensibilité et de la spécificité en fonction des probabilités prédites (Femmes)

***) Analyse statistique des modèles

TAB. 4.15 – Récapitulatif des résultats des estimations : modèle dichotomique de participation au marché du travail

variable dépendante	vardpte = actif occupée			
	Hommes		Femmes	
variables explicatives	coefficients	significativité	coefficients	significativité
constante	4.138	(***)	3.764	(***)
non scolarisé	réf	réf	réf	réf
primaire	0.451	(***)	1.392	(***)
secondaire 1er cycle	0.458	(***)	1.652	(***)
secondaire 2nd cycle	0.519	(**)	2.335	(***)
supérieur	0.750	(***)	3.404	(***)
Bac ou plus	0.534	(***)	0.831	(***)
âge	-0.016	(***)	0.070	(***)
migrant	0.202	(***)	0.082	(-)
situation matrimoniale	0.475	(***)	0.053	(-)
taille ménage	0.056	(***)	0.154	(***)
lien de parenté	1.317	(***)	0.134	(-)
nombre d'enfants en bas age	-0.063	(***)	-0.173	(***)
travaux domestiques	-0.029	(***)	-0.036	(***)
Revenu de réserve	-1.052	(***)	-2.050	(***)
Nombre d'observations	9193		8142	
LR chi2(12)	787.24 (***)		1850.67 (***)	
Pouvoir de classement	88.75		80.18	
Test de Hosmer	1988.70 (0.5544)		2464.76 (0.6783)	
Rapport de vraisemblance : statistique = 66.57 (***)				
(***) = $p_value < 0.001$ (**) = $p_value < 0.01$; (*) = $p_value < 0.05$				
(-) = non significatif				

Interprétation des résultats

Le tableau ci-dessus résume les résultats de l'estimation de l'équation de participation au marché du travail selon le genre. Le test du *chi2* montre que les deux équations sont fortement significatives et le test du rapport de vraisemblance nous permet de valider le fait qu'on ait estimé séparément le modèle selon le genre.

L'observation de ces résultats montre que la participation des femmes ou des hommes au marché du travail est largement influencée par leur niveau d'instruction. En effet, Les coefficients relatifs aux modalités de cette variable sont tous positifs et très fortement significatifs. On note cependant que lorsque l'âge d'un homme augmente d'une unité, sa chance de participer au marché du travail diminue de 1,6%. Mais, ce phénomène est contraire chez la femme qui voit plutôt sa chance augmentée de 7,0% environ.

Les résultats de ce tableau confirme également le fait que l’environnement social d’un individu joue un rôle fondamental dans sa participation à une activité économique. En effet, le fait qu’un homme se marie augmente très significativement sa chance de participer au marché du travail, ce qui n’est pas autant chez les femmes (le coefficient de ce paramètre n’est pas significatif chez elles bien qu’étant positif, équivalent de l’accroissement de la probabilité). De même, le lien qu’il établie avec son chef de ménage a une très grande influence sur sa participation ; cela est cependant plus élevé chez les hommes que les femmes. Le nombre d’heures de travaux domestiques qu’effectue un individu et l’effet du nombre d’enfants en bas âge dans le ménage que celui-ci fait partie, réduisent également très substantiellement ses chances de participer au marché. Le tableau suivant exprime encore plus la portée de ces effets. La colonne dF/dx exprime l’effet marginal d’une variation d’une unité de la variable indépendante sur la probabilité de se retrouver dans un certain état.

TAB. 4.16 – Variation de la probabilité de participation au marché du travail

variable dépendante	vardpte = actif occupée			
	Hommes		Femmes	
variables explicatives	dF/dx	Erreur standard	dF/dx	Erreur standard
non scolarisé	réf	réf	réf	réf
primaire	0.0369932	0.0129128	0.0893843	0.0131239
secondaire 1er cycle	0.034244	0.0109232	0.0733334	0.0084272
secondaire 2nd cycle	0.0343692	0.0080391	0.0528027	0.0045508
supérieur	0.0397523	0.0047415	0.0410914	0.0034518
Bac ou plus	0.0381894	0.007199	0.0433215	0.0059879
âge	-0.0013942	0.0002105	0.0048974	0.0002758
migrant	0.0185688	0.0048633	0.0057861	0.0036788
situation matrimoniale	0.0407977	0.0058462	0.0036788	0.0040772
taille ménage	0.0049473	0.0011867	0.0107477	0.0012179
lien de parenté	0.1741034	0.0133681	0.0097657	0.0054308
nombre d’enfants en bas age	-0.005637	0.0015253	-0.0120282	0.001518
travaux domestiques	-0.0025613	0.0002545	-0.0025049	0.0001793
Revenu de réserve	-0.0935606	0.0055546	-0.1426858	0.0087196

Enfin, chez les hommes comme les chez les femmes, le coefficient du revenu de réserve, significatif et négatif, permet de dire que l’offre de travail n’est pas une fonction croissante du revenu. En d’autres termes, plus le revenu de réserve d’un individu est élevé, moins il est un actant du marché du travail. Cela est d’ailleurs tout à fait normal, car l’on ne peut accepter de travailler, que lorsqu’on gagnerait plus que ce que l’on possède sans travailler.

En conclusion, ces modèles nous ont permis de mettre en évidence certains déterminants clés des décisions de chaque individu de participation au marché du travail. En particulier, ils ont particulièrement révélé l’influence notable des variables du capital humain, notamment le niveau d’instruction. Cela permet en d’autres termes, de dire que l’accent devrait être mis sur la valorisation du secteur éducation. Des résultats de l’analyse descriptive et la mise en évidence de deux fonctions de gains et de participation au marché du travail selon le sexe, nous

permet également de conclure de l'existence d'une sorte de discrimination sur le marché du travail, notamment par le revenu et par l'emploi. L'intérêt de la suite de ce travail consisterait en l'évaluation du degré de discrimination sur le marché du travail.

CONCLUSION

Cette étude avait pour objectif d'analyser le marché du travail au Cameroun. A partir de l'enquête sur l'emploi et le secteur informel réalisée en 2005 par l'Institut National de la Statistique, nous avons essayé d'étudier la décision pour chaque individu, de participer au marché du travail. A partir du modèle probit, nous avons identifié les éléments clés de la participation au marché du travail. Nous avons également estimé par la méthode des moindres carrés ordinaires et suivant le sexe, les gains des actifs sur le marché du travail. Globalement, on a noté l'influence remarquable du niveau d'instruction sur les fonctions de gains et de participation. Les résultats de l'analyse ont permis d'étudier le phénomène de discrimination dans l'emploi occupé à la sortie des individus du système éducatif. L'intérêt a été alors d'entrevoir si le sexe des individus, variable considérée comme sans lien avec la productivité de l'homme, influençait la situation de travail en début et pendant la vie active. Trois aspects de la situation professionnelle ont été abordés : d'une part, les salaires et donc les risques de discrimination salariale, d'autre part, la structure des emplois occupés et son incidence sur les écarts de salaires entre catégories d'emplois et au sein des mêmes catégories, ce selon le sexe des individus, et enfin, l'accès à l'emploi. Pour chacun d'eux, la discrimination a été analysée selon le sexe.

L'analyse de la discrimination de genre révèle un phénomène récurrent dans les différentes mesures de chaque aspect de la situation professionnelle, phénomène particulièrement défavorable à l'encontre des femmes. En effet, celles-ci sont victimes de discrimination salariale de grande envergure. Les écarts salariaux entre hommes et femmes s'expliquent majoritairement par des différences de salaire au sein de mêmes catégories d'emploi, écarts en partie non justifiés par des différences de caractéristiques productives. Notons à ce point que l'un des objectifs de la suite de ce travail, serait d'évaluer le degré de discrimination dans les différents secteurs du marché, particulièrement chez les salariés et les indépendants par la méthode d'Oaxaca telle que présentée à la fin de la section 4.2 de ce travail. On a également noté que l'accès à l'emploi donne également lieu à la discrimination à l'encontre des femmes.

Il reviendrait alors au gouvernement de mettre plus d'accent sur le mode de recrutement qui doit en réalité être basé uniquement sur le mérite, il serait également nécessaire de mettre à la disposition des employeurs, des moyens fiables devant leur permettre de mieux apprécier les acteurs de l'offre. Le gouvernement devrait aussi amener les employeurs à appliquer les lois de la législation du travail, notamment le volet salarial. Il serait aussi très important de prendre des dispositions sévères en l'encontre des acteurs véreux qui tourneraient au revers de leurs doigts ces lois.

ANNEXE

[1] Définition de quelques termes clés utilisés

Une ZD est une portion de territoire délimité par des détails visibles et comprenant 700 à 1100 habitants.

Population active : c'est la population constituée des chômeurs et de la population active occupée.

Population active occupée : c'est la population regroupant l'ensemble des personnes ayant travaillé (ne se reste qu'une heure) au cours de la semaine précédant l'enquête, qu'elles soient salariées, à leur compte, employeurs ou aide familiaux dans l'entreprise ou l'exploitation familiale. Elles comprennent aussi les personnes pourvues d'un emploi, mais qui en sont temporairement absentes pour un motif tel qu'une maladie, des congés payés, un conflit de travail ou une mise en stage.

Population potentiellement active : population en âge de travailler, c'est-à-dire constituée des personnes de 10 ans plus.

[2] Commandes spss d'analyse des données de notre étude

* suppression des isiteurs SEL IF m3 LE 2.

*utilisation du coefficient de pondération wei by coefext3.

***Répartition de la population totale par sexe et par groupe d'âge**

TABLES

/FTOTAL= t "Total"

/TABLE= clage >(STATISTICS) + t BY m4+t

/STATISTICS cpct(m4(F5.1) ' ' :clage)

/ TITLE " Répartition de la population par sexe et par groupe d'âge"

/CAPTION " Source : INS, EESI 2005".

*_____.

TABLES

/FTOTAL= t "Total"

/TABLE= clage (STATISTICS) + t BY m4+t

/STATISTICS cpct(m4(F5.1) ' ' :m4)

/ TITLE " Répartition de la population par sexe et par groupe d'âge"

/CAPTION " Source : INS, EESI 2005".

```
*****
*Répartition de la population totale par sexe et selon le niveau d'instruction
*****
```

```
TABLES
```

```
/FTOTAL= t "Total"
```

```
/TABLE= nivinst ( STATISTICS) + t BY m4+t
```

```
/STATISTICS cpct( m4( F5.1 ) ' ' :nivinst )
```

```
/ TITLE " Répartition de la population totale par niveau d'instruction et par sexe "
```

```
/CAPTION "Source : INS, EESI 2005".
```

```
*_____.
```

```
TABLES
```

```
/FTOTAL= t "Total"
```

```
/TABLE= nivinst ( STATISTICS) + t BY m4+t
```

```
/STATISTICS cpct( m4( F5.1 ) ' ' :m4 )
```

```
/ TITLE " Répartition de la population totale par sexe et par niveau d'instruction "
```

```
/CAPTION " Source : INS, EESI 2005".
```

```
*****
```

```
*——-fin de cette fonction _____
```

```
*****
```

```
*Taux de scolarisation par sexe et par milieu de résidence
```

```
*****
```

```
compute poscol = ( m5 ge 6).
```

```
compute acco2 =100* ( nivinst ge 2).
```

```
compute chobit =100* ( actif = 2).
```

```
filter by poscol.
```

```
TABLES
```

```
/ OBS acco2
```

```
/FTOTAL= t "Total"
```

```
/TABLE= milieu ( STATISTICS) + t BY m4+t by acco2
```

```
/STATISTICS mean( acco2(F5.2 ) )
```

```
/ TITLE " Taux de scolarisation par sexe et par milieu de résidence (urbain ou rural)"
```

```
/CAPTION " Source : INS, EESI 2005".
```

```
*****
```

```
*Taux de scolarisation par sexe et par région
```

```
*****
```

```
TABLES
```

```
/ OBS acco2
```

```
/FTOTAL= t "Total"
```

```
/TABLE= g1 ( STATISTICS) + t BY m4+t by acco2
```

```
/STATISTICS mean( acco2(F5.2 ) )
```

```
/ TITLE " Taux de scolarisation par sexe et par région "
```

/CAPTION " Source : INS, EESI 2005".

filter off.

***Répartition de la population active par sexe et par milieu de résidence**

* p10 = variable qui trie les individus d'âges = à 10 ans ou +.

compute p10 = (m5 ge 10).

compute acco =100* (actif le 2).

compute chobit =100* (actif = 2).

filter by p10.

*_____.

TABLES

/ OBS acco

/FTOTAL= t "Total"

/TABLE= milieu (STATISTICS) + t BY m4+t by acco

/STATISTICS mean(acco(F5.2))

/ TITLE " Taux d'activité par sexe et par milieu de résidence"

/CAPTION " Source : INS, EESI 2005".

***Répartition de la population active par sexe et région**

TABLES

/ OBS acco

/FTOTAL= t "Total"

/TABLE= g1 (STATISTICS) + t BY m4+t by acco

/STATISTICS mean(acco(F5.2))

/ TITLE " Taux d'activité par sexe et par région"

/CAPTION " Source : INS, EESI 2005".

*_____fin de cette fonction_____

*** Taux d'activité de la population par sexe et selon la religion**

TABLES

/OBS acco

/FTOTAL= t "Total"

/TABLE= m8 (STATISTICS) + t BY m4+t by acco

/STATISTICS cpct(m4(F5.2) ' ' :m8)

/ TITLE " Taux d'activité par sexe et par religion"

/CAPTION " Source : INS, EESI 2005".

```

TABLES
/OBS acco
/FTOTAL= t "Total"
/TABLE= m8 ( STATISTICS) + t BY m4+t by acco
/STATISTICS cpct( m4( F5.2 ) ' ' :m4 )
/ TITLE " Taux d'activité par sexe et religion"
/CAPTION " Source : INS, EESI 2005".
*****

compute acco2 =100* ( nivinst ge 2).
TABLES
/ OBS acco2
/FTOTAL= t "Total"
/TABLE= milieu ( STATISTICS) + t BY m4+t by acco2
/STATISTICS mean( acco2(F5.2 ) )
/ TITLE " Taux de scolarisation de la population active par sexe et par milieu de résidence
(urbain ou rural)"
/CAPTION " Source : INS, EESI 2005".
filter off.
*****

*Répartition de la population active selon le niveau d'instruction et par sexe
*****

* à partir de ce point, la population sélectionnée est active par les variables aco3 et p10
compute p10 = (m5 ge 10).
compute acco2 =100* ( nivinst ge 2).
compute acco3 = (( actif le 2) and ( p10 = 1)).
filter by acco3.
TABLES
/FTOTAL= t "Total"
/TABLE= nivinst ( STATISTICS) + t BY m4+t
/STATISTICS cpct( m4( F5.1 ) ' ' :nivinst )
/ TITLE " Répartition de la population active par niveau d'instruction et par sexe "
/CAPTION "Source : INS, EESI 2005".
*_____

TABLES
/FTOTAL= t "Total"
/TABLE= nivinst ( STATISTICS) + t BY m4+t
/STATISTICS cpct( m4( F5.1 ) ' ' :m4 )
/ TITLE " Répartition de la population active par sexe et par niveau d'instruction "
/CAPTION " Source : INS, EESI 2005".
*****

*Taux de scolarisation de la population active selon la région par sexe

```

```

*****
TABLES
/ OBS acco2
/FTOTAL= t "Total"
/TABLE= milieu ( STATISTICS) + t BY m4+t by acco2
/STATISTICS mean( acco2(F5.2 ) )
/ TITLE " Taux de scolarisation de la population active par sexe et par milieu de résidence
(urbain ou rural)"
/CAPTION " Source : INS, EESI 2005".

```

```
*****
```

***Taux de scolarisation de la population active par sexe et par région**

```
*****
```

```

TABLES
/ OBS acco2
/FTOTAL= t "Total"
/TABLE= g1 ( STATISTICS) + t BY m4+t by acco2
/STATISTICS mean( acco2(F5.2 ) )
/ TITLE " Taux de scolarisation de la population active par sexe et par région "
/CAPTION " Source : INS, EESI 2005".

```

filter off.

```
*****
```

***Taux de chômage au sens du BIT par région**

```
*****
```

```

compute p10 = ( m5 ge 10).
compute pchom =((actif = 1 or actif = 2 ) and (p10 = 1) ).
compute chbit =100* ( actif = 2).

```

filter by pchom

```

TABLES
/ OBS chbit
/FTOTAL= t "Total"
/TABLE= g1 ( STATISTICS) + t BY m4+t by chbit
/STATISTICS mean( chbit(F5.2 ))
/ TITLE " Taux de chômage BIT par sexe et par région"
/CAPTION " Source : INS, EESI 2005".

```

filter off.

```
*****
```

***Taux de chômage au sens du BIT par milieu**

```
*****
```

```

compute p10 = ( m5 ge 10).
compute pchom =((actif = 1 or actif = 2 ) and (p10 = 1) ).
compute chbit =100* ( actif = 2).

```

```

filter by pchom
TABLES
/ OBS chbit
/FTOTAL= t "Total"
/TABLE= milieu ( STATISTICS) + t BY m4+t by chbit
/STATISTICS mean( chbit(F5.2 ))
/ TITLE " Taux de chômage BIT par sexe et par région"
/CAPTION " Source : INS, EESI 2005".
filter off.
*****

*Répartition des actifs du secteur formel par sexe et selon l'emploi
*****

compute p10 = ( m5 ge 10).
compute pchom =((actif = 1 or actif = 2 ) and (p10 = 1) ).
compute actfor =100* ( (siind =1) or (siind = 2 ) and (pchom = 1 )).
filter by actfor.
*_____-.

TABLES
/FTOTAL= t "Total"
/TABLE= typempl ( STATISTICS) + t BY m4+t
/STATISTICS cpct( m4( F5.2 ) ' ' :typempl )
/ TITLE " Répartition des actifs du secteur formel par sexe "
/CAPTION "Source : INS, EESI 2005".
*****

TABLES
/FTOTAL= t "Total"
/TABLE= typempl ( STATISTICS) + t BY m4+t
/STATISTICS cpct( m4( F5.2 ) ' ' :m4 )
/ TITLE " Répartition des actifs du secteur formel par sexe "
/CAPTION "Source : INS, EESI 2005".
filter off.
*****

*Répartition des actifs du secteur privé formel par sexe et type d'emploi
*****

compute p10 = ( m5 ge 10).
compute pchom =((actif = 1 or actif = 2 ) and (p10 = 1) ).
compute actfor =100* ( (siind = 2 ) and (pchom = 1 )).
filter by actfor.
*_____-.

TABLES
/FTOTAL= t "Total"

```

```

/TABLE= typempl ( STATISTICS) + t BY m4+t
/STATISTICS cpct( m4( F5.2 ) ' ' :typempl )
/ TITLE " Répartition des actifs du secteur privé formel par sexe "
/CAPTION "Source : INS, EESI 2005".
filter off.
*****

*Répartition des actifs du secteur privé formel par sexe et type d'emploi
*****

compute p10 = ( m5 ge 10).
compute pchom =((actif = 1 or actif = 2 ) and (p10 = 1) ).
compute actfor =100* ( (siind = 2 ) and (pchom = 1) ).
*****

filter by actfor.
*_____

TABLES
/FTOTAL= t "Total"
/TABLE= typempl ( STATISTICS) + t BY g1+t
/STATISTICS cpct(g1( F5.2 ) ' ' :typempl )
/ TITLE " Répartition des actifs du secteur privé formel par sexe "
/CAPTION "Source : INS, EESI 2005".
filter off.
*****

*Répartition des actifs du secteur informel par sexe et selon l'emploi
*****

compute p10 = ( m5 ge 10).
compute pchom =((actif = 1 or actif = 2 ) and (p10 = 1) ).
compute actfor =100* ( (siind =3) or (siind = 4 ) and (pchom = 1) ).
*****

filter by actfor.
TABLES
/FTOTAL= t "Total"
/TABLE= typempl ( STATISTICS) + t BY m4+t
/STATISTICS cpct( m4( F5.2 ) ' ' :typempl )
/ TITLE " Répartition des actifs du secteur informel par sexe "
/CAPTION "Source : INS, EESI 2005".
*****

TABLES
/FTOTAL= t "Total"
/TABLE= typempl ( STATISTICS) + t BY m4+t
/STATISTICS cpct( m4( F5.2 ) ' ' :m4 )
/ TITLE " Répartition des actifs du secteur informel par sexe "

```

```

/CAPTION "Source : INS, EESI 2005".
filter off.
*****

*catégorie socioprofessionnelle des actifs occupés par sexe
*****

compute p10 = ( m5 ge 10).
compute pchom2 =( (actif = 1 ) and (p10 = 1) ).
*****

filter by pchom2.
*_____

TABLES
/FTOTAL= t "Total"
/TABLE= ap3 ( STATISTICS) + t BY m4+t
/STATISTICS cpct( m4( F5.2 ) ' ' :ap3 )
/ TITLE " Catégorie socioprofessionnelle des actifs occupés par sexe "
/CAPTION "Source : INS, EESI 2005".
*****

TABLES
/FTOTAL= t "Total"
/TABLE= ap3 ( STATISTICS) + t BY m4+t
/STATISTICS cpct( m4( F5.2 ) ' ' :m4 )
/ TITLE " Catégorie socioprofessionnelle des actifs occupés par sexe "
/CAPTION "Source : INS, EESI 2005".
filter off.
*****

*Mode de rémunération des actifs occupés par sexe
*****

compute p10 = ( m5 ge 10).
compute pchom2 =( (actif = 1 ) and (p10 = 1) ).
*****

filter by pchom2.
*_____

TABLES
/FTOTAL= t "Total"
/TABLE= ap12 ( STATISTICS) + t BY m4+t
/STATISTICS cpct( m4( F5.2 ) ' ' :ap12 )
/ TITLE " Mode de rémunération des actifs occupés par sexe "
/CAPTION "Source : INS, EESI 2005".
*****

TABLES
/FTOTAL= t "Total"

```

```

/TABLE= ap12 ( STATISTICS) + t BY m4+t
/STATISTICS cpct( m4( F5.2 ) ' ' :m4 )
/ TITLE "Mode de rémunération des actifs occupés par sexe "
/CAPTION "Source : INS, EESI 2005".
filter off.

```

[3] Programme stata des estimations effectuées

```

gen visitor=( m3>=3)
lab var visitor " visiteur"
drop if visitor==1
gen p10= (m5>=10 & actif <=3)
lab var p10 " plus de 10 ans et actif"
sort g1 g3 g5 g6 g7
by g1 g3 g5 g6 g7 : egen np10=sum(p10)
lab var np10 " nombre de personne de plus de 10 ans dans le ménage"
gen situation_matrimoniale=( m6 >= 2 & m6 <=3)
label variable situation_matrimoniale " situation matrimoniale"
gen lien_parente=(m2==1| m2==2)
label variable lien_parente " Chef de ménage ou épouse"
gen nombre_enfants_bas_age= tailm- np10
label variable nombre_enfants_bas_age " Enfant de moins 10 ans dans le menage consi-
déré"
gen t3_5_travailleurs = ( ap5 == 3)
label variable t3_5_travailleurs "la taille de l'entreprise est comprise entre 3 et 5 tra-
vailleurs"
recode t3_5_travailleurs . =0
gen plus_5_travailleurs=( ap5 >= 4)
label variable plus_5_travailleurs" taille de l'entreprise : plus de 5 travailleurs"
recode plus_5_travailleurs . =0
gen paie_impot=( ap6a <= 4)
label variable paie_impot "paiement impôt"
recode paie_impot . =0
gen aumoins_le_bac=( m17 >= 5)
label variable aumoins_le_bac " aumoins le bac"
recode aumoins_le_bac . =0
gen migrant=( natif==2)
label variable migrant " migrant"
gen nationalite=( m7==1)
label variable nationalite "Est camerounais ou pas"
recode nationalite . =0
gen vardpte =( actif==1)
label variable vardpte " variable dépendante du modèle probit"

```

```

gen experience_emploi= ap8a1
label variable experience_emploi " expérience dans l'emploi"
recode experience_emploi (missing=0)
gen experience_potentielle = ap8a2
label variable experience_potentielle " expérience potentielle"
recode experience_potentielle .=0
recode experience_potentielle 99=0
gen taille_menage=tailm
label variable taille_menage "taille du ménage considéré dont fait partir l'individu"
gen travaux_domestiques=se1b
label variable travaux_domestiques " nombre d'heure de travaux domestique que l'individu effectue par semaine"
gen sexe=m4
label variable sexe "sexe de l'individu considéré"
gen Logrevenu = ln( revactp)
label variable Logrevenu " variable dépendante de la modélisation de la fonction de gain"
* creation du filtre pour les actifs occupés
gen filter = (actif==1 )
label variable filter "actifs occupés"
*Modèles permettant d'estimer le ratio de MIills
xi : probit vardpte i.nivinst age migrant situation_matrimoniale aumoins_le_bac
taille_menage lien_parente nombre_enfants_bas_age travaux_domestiques Revenu_h
if sexe==1 & p10==1
gen pred_h=(explicative >= 0.5)
gen bonpred_h =(vardpte-pred_h)
tab bonpred_h if sexe ==1 & p10 == 1
gen pred_h1=(explicative >= 0.913)
gen bonpred_h1 =(vardpte-pred_h1)
tab bonpred_h1 if sexe ==1 & p10 == 1
xi : probit vardpte i.nivinst age migrant situation_matrimoniale aumoins_le_bac
taille_menage lien_parente nombre_enfants_bas_age travaux_domestiques Revenu_h
if sexe==2 & p10==1
lfit, group(2500)
lsens
lstat, cutoff(0.8)
lroc
gen pred_f=(explicative >= 0.8)
gen bonpred_f =(vardpte-pred_f)
tab bonpred_f if sexe ==2 & p10 == 1
*****
gen pred_f2=(explicative >= 0.5)

```

```

gen bonpred_f2 =(vardpte-pred_f2)
tab bonpred_f2 if sexe ==2 & p10 == 1
xi : probit vardpte i.nivinst age migrant situation _matrimoniale aumoins _le _bac
taille _menage lien _parente nombre _enfants _bas _age travaux _domestiques if sexe==1
& p10==1, nolog
predict explicative if sexe==1 & p10==1,xb
gen pred1=(explicative >= 0.5)
gen bonpred1 =(vardpte-pred1)
tab bonpred1 if sexe ==1 & p10 == 1
xi : probit vardpte i.nivinst age migrant situation _matrimoniale aumoins _le _bac
taille _menage lien _parente nombre _enfants _bas _age travaux _domestiques if sexe==2
& p10==1, nolog
predict explicative1 if sexe==2 & p10==1, xb
gen pred2=(explicative1 >= 0.5)
gen bonpred2 =(vardpte-pred2)
tab bonpred2 if sexe ==2 & p10 == 1
replace explicative = explicative1 if sexe==2 & p10==1
gen inv_mills_hf = normden(explicative) / norm(explicative)
***** Fonction de gain des HOMMES
xi : reg Logrevenu i.nivinst experience _emploi experience _potentielle migrant nationalite
t3_5_travailleurs plus_5_travailleurs paie _impot inv_mills_hf if sexe==1 & filter==1
& p10==1
*predict expli_reve_h if sexe==1 & p10==1, xb
*****
****Analyse des résidus
predict residu11 if sexe==1 & p10==1, resid
tway (scatter residu11 Revenu_h if sexe==1 & p10==1)
gen individu = _n
tway (scatter residu11 individu if sexe==1 & p10==1)
*list individu if residu11<-2
*list individu if residu11> 2
*drop if ( (residu11 <-2 | residu11>2) & (sexe==1 & p10==1) )
* Fonction de gain définitive
xi : reg Logrevenu i.nivinst experience _emploi experience _potentielle migrant nationalite
t3_5_travailleurs plus_5_travailleurs paie _impot inv_mills_hf if sexe==1 & filter==1
& p10==1 & (residu11 >= -2 & residu11 <= 2)
predict residu12 if sexe==1 & p10==1, resid
predict Revenu_h if sexe==1 & p10==1, xb
tway (scatter residu12 Revenu_h if sexe==1 & p10==1)
tway (scatter residu12 individu if sexe==1 & p10==1)
***** Fonction de gain des FEMMES

```

```

xi : reg Logrevenu i.nivinst experience_emploi experience_potentielle migrant nationalite
t3_5_travailleurs plus_5_travailleurs paie_impot inv_mills_hf if sexe==2 & filter==1
& p10==1
****Analyse des résidus
predict residu21 if sexe==2 & p10==1, resid
tway (scatter residu21 individu )
* list id if residu21<-2
*list id if residu21> 2
*drop if ( (residu21 < -2 | residu21 > 2) & (sexe==2 & p10==1))
* Fonction de gain définitive
xi : reg Logrevenu i.nivinst experience_emploi experience_potentielle migrant nationalite
t3_5_travailleurs plus_5_travailleurs paie_impot inv_mills_hf if sexe==2 & filter==1
& p10==1 & (residu21 >= -2 & residu21 <= 2)
predict residu22 if sexe ==2 & p10==1, resid
predict Revenu_f if sexe==2 & p10==1,xb
tway (scatter residu21 Revenu_f )
tway (scatter residu22 individu )
replace Revenu_h = Revenu_f if sexe ==2 & p10==1
*****

* estimation finale des équations de participation sur le marché de travail
xi : probit vardpte i.nivinst age migrant situation_matrimoniale aumoins_le_bac
taille_menage lien_parente nombre_enfants_bas_age travaux_domestiques Revenu_h
if sexe==1 & p10==1
gen pred_h=(explicative >= 0.5)
gen bonpred_h =(vardpte-pred_h)
tab bonpred_h if sexe ==1 & p10 == 1
gen pred_h1=(explicative >= 0.913)
gen bonpred_h1 =(vardpte-pred_h1)
tab bonpred_h1 if sexe ==1 & p10 == 1
xi : probit vardpte i.nivinst age migrant situation_matrimoniale aumoins_le_bac
taille_menage lien_parente nombre_enfants_bas_age travaux_domestiques Revenu_h
if sexe==2 & p10==1
lfit, group(2500)
lsens
lstat, cutoff(0.8)
lroc
gen pred_f=(explicative >= 0.8)
gen bonpred_f =(vardpte-pred_f)
tab bonpred_f if sexe ==2 & p10 == 1
*****

gen pred_f2=(explicative >= 0.5)
gen bonpred_f2 =(vardpte-pred_f2)

```

```
tab bonpred_f2 if sexe ==2 & p10 == 1
xi : probit vardpte i.nivinst age migrant nationalite situation _matrimoniale aumoins _le _bac
taille _menage lien _parente nombre _enfants _bas _age travaux _domestiques Revenu _h
if p10==1
* estimation des effets marginaux de participation sur le marché de travail
xi : dprobit vardpte i.nivinst age migrant situation _matrimoniale aumoins _le _bac
taille _menage lien _parente nombre _enfants _bas _age travaux _domestiques Revenu _h
if sexe==1 & p10==1
xi : dprobit vardpte i.nivinst age migrant situation _matrimoniale aumoins _le _bac
taille _menage lien _parente nombre _enfants _bas _age travaux _domestiques Revenu _h
if sexe==2 & p10==1
xi : dprobit vardpte i.nivinst age migrant situation _matrimoniale aumoins _le _bac
taille _menage lien _parente nombre _enfants _bas _age travaux _domestiques Revenu _h
if p10==1
```

Médiagraphie

- [1] J.P. Lachaud ; les femmes et les marchés du travail en Afrique Subsaharienne
- [2] BIT ; Indicateurs clés du marché du travail.
- [3] Aka Kouamé et Donatien Tameko ; Inégalité entre hommes et femmes sur le marché du travail ; Juin 2000. 20p.
- [4] François Combarrous ; Discrimination du marché du travail urbain : le cas d'Abidjan en Côte d'Ivoire ; 49p.
- [5] J.P. Lachaud ; fondements des modèles à choix qualitatifs ; 1996.
- [6] P. Tassi ; Méthodes Statistiques ; Collection Economica ; 2^{ème} édition ; 1989.
- [7] C. Hurlin ; Econométrie des variables qualitatives, Janvier 2003.
- [8] Manuel de concepts et de définitions utilisés dans les publications statistiques au Cameroun ; Mai 2005 ; 4^{ème} édition ; 179p..
- [9] Samuel Inack Inack ; Economie du capital Humain ; Notions d'Economie du travail ; 3^{ème} Année de Licence.
- [10] Les statistiques sur l'emploi et le marché du travail au Cameroun ; Décembre 2004.
- [11] Dogmo Tatong Clovis ; Dynamique d'une horticulture périurbaine : Exemple du sud Cameroun entre 1995 et 2004 ; Mémoire de Master 2 de Statistique Appliquée ; 2004/2005 ; 154p.
- [12] Jean-Philippe El Aynaoui, ATER ; Participation, choix occupationnel et gains sur un marché du travail segmenté : une analyse appliquée au cas du Maroc.
- [13] J. Scott Long, Regressions Models for Categorical and Limited Variables.