

TABLE DES MATIERES

DEDICACE.....	ivv
REMERCIEMENTS	v
RESUME	viii
ABSTRACT	viii
ABREVIATION	ix
INTRODUCTION.....	1
CHAPITRE 1 : PRESENTATION DE LA STRUCTURE D'ACCUEIL	3
1.1. ORGANISATION DE L'INS	3
1.2. LES MISSIONS DE L'INS.....	4
1.3. LA SOUS DIRECTION DES ÉTUDES ET DE LA NORMALISATION STATISTIQUE	4
CHAPITRE 2 : REVUE DE LA LITTERATURE	6
2.1. CONCEPT DE SECTEUR INFORMEL	6
2.1.1. Définition et historique.....	6
2.1.2. Caractéristiques générales	8
2.2. CARACTERISTIQUES DU SECTEUR INFORMEL AU CAMEROUN.....	9
2.2.1. La pauvreté : définition du concept et mesure	9
2.2.1.1. Définition	9
2.2.2. Mesures de la pauvreté	10
2.2.2.1. La mesure par le revenu.....	11
2.2.2.2. L'approche par les conditions d'existence	11
2.2.5. L'approche subjective	12

CHAPITRE 3 : DEMARCHE METHODOLOGIQUE ET PRESENTATION DU	
MODELE DE REGRESSION UTILISE.....	13
3.1. CHOIX D'UN SEUIL DE PAUVRETE.....	13
3.2. CONSTRUCTION DES SCORES PRIVATIFS	13
3.2.1. Sélections des commodités.....	14
3.2.2. Codage des variables qualitatives.....	15
Variables	15
3.2.3. Définition des scores privatifs.....	16
3.2.4. Seuil de pauvreté de qualité de vie.....	16
3.2.5. Quelles sont les conditions de vie par secteur d'activité ?	18
3.3. METHODES STATISTIQUES.....	18
3.3.1. Choix des variables d'étude.....	19
3.3.2. Outils théoriques du modèle logit	21
3.3.2.1. Rappel sur le modèle de régression linéaire	22
3.3.2.2. Nécessité d'un autre modèle de régression : le modèle logistique	22
3.3.2.3. Définition du modèle Logit	23
3.3.2.4. Importance du modèle Logit : l'Odds Ratio	23
3.3.2.5. Estimation du modèle	24
3.3.3. Evaluation et validation du modèle.....	27
3.3.3.1. Test des paramètres du modèle.....	27
3.3.3.2. Test de Wald	27
3.3.3.3. Test de Rapport de Vraisemblance	27
3.3.3.4. L'adéquation du modèle aux données « Goodness of fit ».....	28
3.3.3.5. Evaluation de la calibration du modèle : le test de Hosmer et Lemeshow	29
3.3.3.6. L'analyse des résidus	30
3.3.3.7. Le Résidu de Pearson est donné par la formule suivante :	31
3.3.4. Evaluation du pouvoir discriminant du modèle : sensibilité, spécificité et courbe ROC	31

CHAPITRE 4: PRÉSENTATION DES RESULTATS	33
4.1. ESTIMATION	33
4.2. RESULTAT DE L'ESTIMATION.....	34
4.2.1. Adéquation du modèle.....	35
4.2.3. Analyse des résidus	37
4.3. INTERPRETATION DES RESULTATS.....	38
CHAPITRE 5: CONCLUSION ET RECOMMANDATIONS	40
BIBLIOGRAPHIE.....	42

DEDICACE

*Je dédie
ce mémoire à tous les
membres de ma famille
en particulier à mon regretté père.*

REMERCIEMENTS

Qu'il me soit permis à la fin de ce modeste travail de remercier:

Le Pr. Henri GWET, pour avoir accepté de superviser ce travail. La motivation et les conseils qu'il a su nous apporter durant la formation et pendant la rédaction de ce mémoire ont été d'un soutien immesurable dans l'accomplissement de ce travail.

Le Dr. Eugène, NDONG NGUEMA qui a accepté le sujet de ce mémoire et a suivi sa réalisation avec beaucoup d'intérêt. Sa disponibilité, son dynamisme et ses judicieux conseils m'ont été d'une inestimable aide dans l'accomplissement de ce travail.

M. Arouna ANJUENNEYA NJOYA, qui m'a confié le sujet, a mis à ma disposition les données et toutes les informations indispensables à sa compréhension, et à son encadrement dans la production de ce mémoire.

Tous les enseignants du Master de Statistique Appliquée pour les ressources matérielles qu'ils ont rendu disponibles et les enseignements qu'ils nous ont dispensés, assurant ainsi notre initiation à la recherche.

Mes remerciements vont également aux personnes suivantes :

Messieurs F. Mano et P. TaKam pour tous leurs conseils

Esther Kenfack, Florence Magnifo, Alain Beka.

Mes amis M. Fonkoua, O. Farikou, C. Etonga, O.Abdoulaye, B.Hema, F. Fotso, Aboubakary, A. Ngock, J. B.Nankap, S. Youmbi, M. Meye, M. Djinang, N. Ndjondo, O. Manga, D. Tchappe, H. Djoufack, C. Zoua, P.Fouelefack, Vachala, S.Tcheutchoua, O. Kamta,

Mes camarades de promotion en particulier A. Hyefouais, H. Woubou et A. Ngoko.

Tous les membres de la Calybib.

Mes remerciements vont enfin à l'endroit de tous ceux qui de près ou de loin ont contribué à la réalisation de ce travail.

RESUME

Actuellement, la réduction de la pauvreté est l'un des deux défis majeurs de la communauté internationale; l'autre défi majeur est d'ordre climatique : la réduction des gaz à effet de serre de manière à stopper et à réduire le réchauffement de la terre. Cette volonté se matérialise à travers les objectifs du millénaire pour le développement (OMD) tels que la réduction de moitié de l'extrême pauvreté (population dont le revenu est inférieur à un dollar par jour) et la faim entre 1990 et l'an 2015. Au-delà des ces objectifs de court et de moyen terme, le but est de réduire la pauvreté de manière durable, en particulier dans les pays en développement qui sont les plus affectés.

Ce mémoire traite de la réduction de la pauvreté chez les acteurs du secteur informel au Cameroun. Il s'articule autour de trois principaux axes : le premier fait état de la construction des scores de privation ; il nous a permis de choisir un seuil de pauvreté et d'identifier les pauvres. Nous avons, par exemple, constaté que le secteur informel avec 55,5% de pauvres, était le secteur institutionnel le plus menacé par la pauvreté. Le second axe quant à lui identifie les déterminants de la pauvreté chez les acteurs du secteur informel au Cameroun ; cette identification s'est faite à l'aide de la modélisation logistique. Nous avons vu par exemple que la catégorie socio professionnelle et le niveau d'instruction ont une influence significative sur le niveau de vie des individus exerçant dans le secteur informel. Le troisième axe propose des variables sur lesquelles il faudrait intervenir pour améliorer le niveau de vie des opérateurs du secteur informel.

ABSTRACT

Presently, poverty reduction is one of the major challenges of the international community; the other challenge is of climatic nature, reduction of greenhouse in a way to stop and reduce the effects of global warming. This commitment is shown through the millennium objectives for development (MOD) such as the reduction by half of extreme poverty (that is that section of the population whose revenue is less than 1 dollar a day) and hunger between 1990 and the year 2015. Apart from these objectives which are in the short and medium terms, the goal is to reduce poverty in a sustainable way, more particularly in developing countries which are the most affected.

This thesis treats poverty reduction within actors of the informal sector in Cameroon. It revolves around three main axes: the first part shows the construction of deprivation; it has helped us to select a poverty level and identify the poor. We have for example discovered that the informal sector, with 55% of the poor, was the most fragile institutional sector as far as poverty is concerned. The second axis identifies poverty determinants of actors of the informal sector in Cameroon; this identification was done with the help of model sing logistics. We saw for example that the socio-professional category and the level of education have a very significant influence on the living standards of the people who worked in the informal sector. The third axis proposes variables upon which we have to intervene in order to ameliorate the living standards of those people in the informal sector.

RESUME EXECUTIF

Ce mémoire traite de la réduction de la pauvreté chez les acteurs du secteur informel au Cameroun. Il nous paraît essentiel de signaler que les données utilisées dans notre étude sont issues de l'EESI phase 1. L'EESI est une enquête statistique à deux phases visant à évaluer la situation de l'emploi (phase 1) et les activités économiques des ménages dans le secteur informel (phase 2).

La base de données de la phase 1 comporte **38 599** individus pour **234** variables. La variable qui a retenu notre attention est la variable secteur institutionnel (SIIND) et un individu dont le secteur institutionnel est le secteur informel (agricole et non agricole) est considéré comme une unité statistique pour notre étude.

Dans le rapport principal de l'EESI phase 1, on pouvait déjà noter que le secteur informel, avec 90,4 % des actifs occupés (dont 55,2 % dans le secteur agricole), fournit le plus d'opportunités d'insertion économique malgré les conditions de travail très précaires.

Puisque dans la base de données dont nous disposons, aucune variable ne pouvait a priori indiquer si un individu était pauvre ou non, il nous a paru utile de construire une variable que nous avons nommée « niveau de vie » prenant deux modalités 1 si l'individu était pauvre et 0 sinon. Ainsi, sur la base de la revue de la littérature que nous avons faite au préalable et sur la base de notre bon sens propre, nous avons construit des scores de privations et nous avons choisi un seuil de pauvreté des conditions d'existence de manière à isoler une incidence de la pauvreté du même ordre que la pauvreté monétaire.

La construction de ces scores privatifs nous a permis de constater que le secteur informel, avec 55,5 % de pauvres, était le secteur institutionnel le plus menacé par la pauvreté des conditions d'existence.

L'objectif général de notre étude était de déterminer les variables sur lesquelles il faudrait agir pour réduire la pauvreté chez les acteurs du secteur informel. Pour atteindre cet objectif, nous avons utilisé le Modèle Logistique et nous avons montré

que les variables sur lesquelles il faudrait agir sont : le niveau d'instruction, la catégorie socioprofessionnelle, le milieu de résidence et l'âge. On peut noter que : les individus ayant un niveau d'instruction élevé (supérieur) présentent moins de risque d'être pauvre contrairement à ceux ayant un niveau d'instruction faible (primaire et moins). On peut également noter que les acteurs du secteur informel vivant en milieu rural et semi urbain présentent plus de risques d'être pauvre que ceux vivant en milieu urbain ; la catégorie socioprofessionnelle laisse présager une vulnérabilité à la pauvreté très élevée chez les manœuvres, les apprentis et chez les ouvriers non qualifiés par rapport aux cadres supérieurs.

Pour pouvoir améliorer les conditions de vie des acteurs du secteur informel, l'Etat devrait :

- Développer les interventions visant à améliorer l'accès des actifs informels à l'éducation ;
- Favoriser le développement des activités économiques en milieu rural et semi urbain ;
- Mettre la pression sur les entreprises pour qu'elles donnent le minimum nécessaire aux ouvriers afin qu'ils puissent avoir des conditions de vie acceptables.

ABREVIATION

BIT :	Bureau International du Travail
BUCREP :	Bureau Central du Recensement et des Etudes de la Population
CIST :	Conseil International des Statistiques du Travail
DEM :	Division des Enquêtes auprès des Ménages
DSCN :	Direction de la Statistique et de la Comptabilité Nationale
ECAM :	Enquête Camerounaise Auprès des Ménages
EESI :	Enquête sur l'Emploi et le Secteur Informel
ENS :	Etudes et Normalisation Statistique
INS:	Institut National de la Statistique
MINPLADAT :	Ministère de la Planification, la Programmation, du Développement et de l'Aménagement de Territoire.
OMD :	Objectifs du Millénaire pour le Développement
PNUD :	Programme des Nations Unies pour le Développement
RGPH :	Recensement Général de la Population et de l'Habitat
UPI :	Unité de Production Informelle
ZD :	Zone de Dénombrement

INTRODUCTION

Depuis plus d'une décennie, la lutte contre la pauvreté est devenue l'axe principal des politiques macro-économiques africaines. Cette volonté de réduire la pauvreté n'est cependant pas récente. En Afrique subsaharienne, au sortir des indépendances, tous les efforts ont été orientés vers la croissance et la prospérité des peuples. Mais ces efforts n'ont pas porté les fruits attendus puisqu'un quart de siècle plus tard, c'est-à-dire vers la fin de la décennie 80, presque tous les pays de cette zone ont connu une période de crise économique ; cela a entraîné une augmentation notable du nombre de pauvres. Dans son rapport sur le développement dans le monde publié en 2000/2001, « Combattre la pauvreté », la Banque mondiale dénombrait déjà 26,6% de pauvres en 1987 et 27,6%¹ en 1990 en Afrique subsaharienne.

Dans ce même rapport, la Banque Mondiale présente trois stratégies de lutte contre la pauvreté à savoir :

- le développement des opportunités ;
- l'insertion ;
- la sécurité matérielle.

Au Cameroun, les principales opportunités se trouvent dans le secteur informel qui s'est développé au cours des années. Les estimations actuelles attestent que plus de 90% des travailleurs se trouvent dans ce secteur d'activité². Ce développement du secteur informel semble se faire avec une paupérisation croissante des populations et notamment des actifs informels. En 2001, la deuxième enquête camerounaise auprès des ménages a montré que les acteurs du secteur informel (agricole et non agricole) étaient les plus pauvres.³

Face à cette situation, les questions suivantes s'imposent :

- Quelle est la situation des actifs informels ?
- Quels sont les déterminants de la pauvreté des actifs informels ?
- Que faut-il faire pour réduire la pauvreté des actifs informels ?

¹ Suivant l'approche monétaire de la pauvreté.

² Rapport sur l'emploi et le secteur informel au Cameroun en 2005. Phase 1 : enquête sur l'emploi. Rapport principal, INS, 2005.

³ Conditions de vie des populations et profil de pauvreté au Cameroun en 2001, ECAM II, INS, 2001

L'objectif général de cette étude est de déterminer les variables sur lesquelles on peut agir pour réduire la pauvreté des acteurs du secteur informel au Cameroun pour faciliter la prise de décisions visant à réduire la pauvreté de manière durable. Les deux objectifs spécifiques sont les suivants :

- 1- Mesurer, analyser et caractériser la situation des acteurs du secteur informel par rapport à la pauvreté ;
- 2- Proposer des axes d'amélioration de leurs activités pouvant jouer sur leur niveau de vie.

Le présent document est divisé en cinq chapitres. Le premier chapitre présente l'INS, structure dans laquelle nous avons effectué notre stage, le deuxième fait état d'une revue de la littérature sur les concepts de pauvreté et de secteur informel ; le troisième fait état de la méthodologie d'analyse ; le quatrième chapitre présente les résultats et les analyses. Les conclusions et les recommandations de l'étude sont présentées dans le dernier chapitre.

CHAPITRE 1 : PRESENTATION DE LA STRUCTURE D'ACCUEIL

Nous avons effectué notre stage à l'Institut National de la statistique du Cameroun. Nous avons simultanément mené des travaux auprès de la Division des Enquêtes et Études Statistiques auprès des Ménages (DEM) et à la Sous Direction des Études et de la Normalisation Statistique (ENS).

Nous présentons l'INS dans son organisation, les missions qui lui sont assignées, ensuite nous présentons la Sous Direction des Études et de la Normalisation Statistique, Sous Direction dans laquelle nous avons été affectés à notre arrivée.

1.1. ORGANISATION DE L'INS

L'Institut National de la Statistique du Cameroun est un établissement public administratif, doté de la personnalité juridique et de l'autonomie financière. Il naît en lieu et place de la Direction de la Statistique et de la Comptabilité Nationale par décret présidentiel N°2001/100 du 20 avril 2001. L'INS est sous tutelle technique du Ministère de la Planification, la Programmation, du Développement et de l'Aménagement de Territoire. L'INS fonctionne encore selon l'organigramme de la DSCN qui comprend donc :

- la Direction Générale ;
- le Service des Affaires Générales ;
- le Service d'ordre ;
- la Division des Enquêtes et des Études Statistiques auprès des Ménages ;
- la Sous Direction des Études et de la Normalisation Statistiques ;
- la Sous Direction de la Comptabilité Nationale et des Synthèses Statistiques ;
- la Sous Direction des Statistiques d'Entreprises ;
- la Sous Direction des Statistiques Démographiques et sociales ;
- la Sous Direction de l'informatique et des banques des données ;
- le Secrétariat permanent au plan comptable ;
- des Services Extérieurs dans les dix Chefs lieux de province.

1.2. LES MISSIONS DE L'INS

Les principales missions confiées à l'INS sont :

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

1.3. LA SOUS DIRECTION DES ÉTUDES ET DE LA NORMALISATION STATISTIQUE

La Sous Direction des Études et de la Normalisation Statistique est placée sous l'autorité d'un Sous Directeur et est chargée :

- du développement des méthodes, des enquêtes et autres études statistiques ;
- des études relatives à l'adéquation de la production nationale de statistiques aux besoins des différentes catégories d'utilisateurs ;
- de l'actualisation de la réglementation statistique ;
- de la centralisation de l'information sur les programmes et projets statistiques des administrations et organismes publics ;
- de l'instruction des dossiers de demande de visas statistiques ;
- de la préparation des dossiers et de tous les éléments concourant au bon déroulement des sessions du conseil national de la statistique ;
- de la centralisation des informations sur la conjoncture ;
- de la publication et de la diffusion des statistiques de l'institut ;
- de la gestion des fonds documentaires et des archives de la direction.

Elle comprend :

- le Service des Études Statistiques ;
- le Service de la Normalisation Statistique ;
- le Service des Publications, de la Diffusion, de la Communication et des Archives.

Le Service des Études Statistiques a été notre service d'accueil. Il est placé sous l'autorité d'un chef de service et est chargé :

- du développement, avec les services concernés, 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'exécution de ces plans ;
- de la préparation, en liaison avec les autres services de l'institut, des projets de programmation annuels et pluriannuels de 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'information statistiques, notamment sur le plan de la méthodologie ;
- de l'instruction des dossiers de demande de visas statistiques.

Le Service des Études Statistiques comprend :

- le Bureau des Études et Méthodes ;
- le Bureau des Programmes.

CHAPITRE 2 : REVUE DE LA LITTERATURE

Ce chapitre présente des éléments de définitions et de discussions tirés de la littérature récente sur la pauvreté et le secteur informel. Il se divise en deux sections. La première est consacrée au concept de secteur informel, la deuxième section traite des aspects liés à la mesure de la pauvreté.

2.1. CONCEPT DE SECTEUR INFORMEL

2.1.1. Définition et historique

Dans la littérature économique, l'expression « secteur informel » est apparue pour la première fois en Afrique en 1965 au Ghana dans une étude portant sur l'emploi urbain.

Pour cerner les réalités du secteur informel, beaucoup de chercheurs ont recours à de nouvelles dénominations pour mieux qualifier le dit secteur à la loupe de leurs préoccupations; et c'est pourquoi, il n'est pas rare de rencontrer des terminologies comme secteur de transition, activités marginales, secteur non structuré, activités tradi-modernes, économie populaire, etc.

A partir des caractéristiques générales issues d'enquêtes antérieures, une définition internationale du secteur informel a pu être adoptée en 1993 au 15^{ème} Conseil International des Statisticiens du Travail, et a permis d'établir des évaluations dans les pays où ce secteur joue un grand rôle en termes d'emplois et de contribution au PIB.

On peut alors retenir comme définition du secteur informel « l'ensemble constitué d'unités économiques produisant des biens et des services en vue de créer principalement des emplois et des revenus, travaillant à petite échelle, avec un faible niveau d'organisation et une faible division entre capital et travail. Les relations de travail recouvrent étroitement les relations de parenté, personnelles et sociales ».

Selon Jacques Charmes⁴, trois remarques doivent accompagner cette définition :

⁴ Auditions : Définition et Evolution du Secteur Informel.

- Ces petites unités de production se caractérisent avant tout par leur non-enregistrement ou celui des travailleurs, et ce, souvent par défaut ou incapacité des systèmes administratifs. Cependant, le secteur informel, qui concerne des activités dites « normales » et ne cherchant pas à se cacher se différencie de l'économie souterraine. Alors que cette économie fait référence au travail illégal et à la fraude fiscale, les opérateurs du secteur informel ne cherchent pas délibérément à échapper aux réglementations.

- Une autre caractéristique importante de cette définition adoptée en 1993 est sa distinction entre les activités agricoles et les activités informelles qui ne sont pas agricoles. Cette distinction est opérée sur des bases empiriques et non théoriques.

- Enfin, autre remarque importante, le concept a souvent été critiqué pour son dualisme. Le terme secteur laisse à penser que deux secteurs sont strictement séparés et qu'il n'y a pas de relation entre les deux. Mais l'adoption de ce terme traduit la volonté d'intégrer le concept dans les systèmes de comptabilité nationale. C'est pourquoi le secteur informel a été défini à partir des caractéristiques de l'unité économique. Le secteur informel est donc selon cette définition un sous-ensemble du secteur institutionnel des ménages en comptabilité nationale.

Dans les années 70, le secteur informel a pris de l'ampleur au point de concurrencer le secteur formel. Déjà à l'époque coloniale, les agents économiques arrivaient à soustraire leur production de l'impôt dit "impôt de capitation", dénoncé par les leaders des mouvements de résistance.

Le développement du chômage urbain, conséquence logique de la crise économique des années 80 et 90, s'est accompagné de l'émergence et de l'essor du secteur informel. C'est une question de survie des populations qui n'arrivent pas à s'intégrer dans le secteur formel. Le secteur informel joue un rôle d'adoption des migrants et un rôle d'accueil des agents économiques exclus du secteur officiel. C'est une zone tampon entre le secteur traditionnel rural (le monde paysan) et le secteur moderne.

La fin du troc a chassé la population rurale vers le secteur informel, faute d'être captée par la fonction publique. La divisibilité des produits, et leur demande évolutive en raison de la faiblesse de leur pouvoir d'achat, est un facteur qui renforce le couple

vendeur/acheteur. Cette relation au-delà de sa fonction économique, favorise une convivialité absente dans le secteur formel où les prix ne donnent en général pas lieu à des négociations.

2.1.2. Caractéristiques générales

Dans les études empiriques, les premières caractéristiques des activités informelles ont vu le jour au Kenya en 1972. Après une étude réalisée sur ce secteur, le BIT en donne les caractéristiques suivantes :

- la facilité d'accès aux métiers ;
- le recours aux ressources locales ;
- la propriété familiale de l'entreprise ;
- l'échelle restreinte d'opérations ;
- l'utilisation des techniques à forte intensité de main-d'œuvre et adaptation au milieu ;
- l'acquisition des qualifications en dehors du système scolaire officiel ;
- la facilité d'opérer sur des marchés non réglementés, mais ouverts et compétitifs⁵.

Sur la base de la complexité des activités du secteur informel et d'autres études à travers le monde, d'autres caractéristiques et critères se sont ajoutés. Signalons ceux proposés par Ilenda⁶ :

- ne pas bénéficier d'un crédit bancaire ;
- présenter un caractère provisoire ou ambulant ;
- se contenter des investissements faibles ;
- ne pas tenir de comptabilité ;
- ne pas inscrire le personnel à la caisse de la sécurité sociale ;
- avoir un horaire de travail irrégulier ;
- ne pas avoir de statut juridique.

⁵ BIT, Rapport sur les stratégies pour accroître l'emploi productif au Kenya, Genève, BIT, 1974, p42.

⁶ Banque mondiale, Zaïre : Orientations stratégiques pour la reconstruction économique, Washington, Banque mondiale, novembre 1994, p.2-3 inédit.

2.2. CARACTERISTIQUES DU SECTEUR INFORMEL AU CAMEROUN

La littérature sur le secteur informel au Cameroun est rare. C'est en 1993 qu'une première étude sur ce secteur d'activité a été réalisée dans la capitale camerounaise Yaoundé. Cette étude avait relevé la multiplication des créations d'Unités de Production Informelles; ce qui se traduisait par la montée en puissance du secteur informel comme mode d'insertion privilégié de la main-d'œuvre en période de crise.⁷ Elle avait également mentionné plusieurs caractéristiques du secteur informel au Cameroun :

- une précarité des conditions d'activité. Près de 80% des UPI sont abritées par des installations de fortune ; elles sont en général méconnues des services de l'État et 16% des UPI payent la patente ;
- absence générale de protection sociale ;
- exclusion du système bancaire des UPI ;

Il faut attendre Décembre 2005⁸ pour avoir les caractéristiques du secteur informel sur le plan national :

- ses activités sont concentrées en milieu Urbain ;
- faible scolarisation des opérateurs ;
- la main d'œuvre est jeune ;
- la moitié des actifs informels sont des femmes ;
- le revenu mensuel des travailleurs est faible.

2.2.1. La pauvreté : définition du concept et mesure

2.2.1.1. Définition

Depuis le début des années quatre-vingt dix, la pauvreté est un thème récurrent au sein des organisations internationales de développement. En 1990, pour la première fois, la Banque Mondiale consacre explicitement le rapport sur le développement dans

⁷ Il faut signaler qu'au moment où se réalisait cette étude le Cameroun subissait une récession économique.

⁸ Enquête sur l'Emploi et le Secteur Informel au Cameroun en 2005. Phase I : Enquête sur l'emploi rapport principal

le monde à cette notion. De son côté, le PNUD porte une attention croissante à cette réalité sociale. Après avoir élaboré la notion de développement humain, le PNUD avance celle de pauvreté humaine.

Aujourd'hui, ces deux organismes s'accordent sur le fait que la pauvreté est un phénomène complexe, pluridimensionnel ne pouvant être réduit à sa simple expression monétaire.

Le PNUD, dans son rapport « Vaincre la Pauvreté Humaine » publié en 2000, présente une définition de « l'extrême pauvreté », de la « pauvreté générale » et la « pauvreté humaine ». Ainsi, une personne vit dans la pauvreté extrême si elle ne dispose pas de revenus nécessaires pour satisfaire ses besoins alimentaires essentiels. Une personne vit dans la pauvreté générale si elle ne dispose pas de revenus suffisants pour satisfaire ses besoins essentiels non alimentaires. La pauvreté humaine est quant à elle l'absence des capacités humaines de base.

Les types de pauvreté abordés par la Banque Mondiale sont centrés sur l'aspect monétaire. Elle distingue pauvreté absolue et pauvreté relative. La pauvreté absolue concerne les personnes dont le niveau de revenu est inférieur à celui nécessaire pour assurer la survie. La pauvreté relative, quant à elle, reflète une conception plus axée sur la répartition des revenus. Elle concerne les personnes situées au plus bas de l'échelle des revenus.

Pour ce faire une idée un peu plus précise sur la manière dont la pauvreté est appréhendée, nous présentons dans la suite les indicateurs utilisés pour la mesurer.

2.2.2. Mesures de la pauvreté

La grande majorité des travaux existant sur le sujet se sont surtout interrogés sur le ou les meilleurs critères à retenir pour définir le concept de pauvreté et les difficultés pour la mesurer.

Les analystes de la Banque Mondiale tendent à privilégier une évaluation monétaire et ceux du PNUD font une évaluation plus globale par les conditions d'existence.

2.2.2.1. La mesure par le revenu

Parmi les multiples approches de mesure de la pauvreté, celles qui s'intéressent à l'aspect monétaire sont les plus fréquentes : elles considèrent comme pauvre, toute personne dont les revenus sont inférieures à un certain seuil dit seuil de pauvreté. Ces seuils peuvent être des seuils absolus ou des seuils relatifs. Les seuils absolus font référence à l'idée de «minimum vital» par rapport au niveau et au mode de vie qui prévalent dans une société à un moment donné.

Les mesures de la pauvreté fondées sur le revenu sont objectives, se prêtent très bien à une analyse quantitative et décrivent avec précision la pauvreté monétaire. Cependant, elles omettent les facteurs autres que monétaires, comme les niveaux sanitaires et éducatifs.

2.2.2.2. L'approche par les conditions d'existence

C'est une approche de la pauvreté en termes de privation relative qui a d'abord été développée par Peter Townsend puis, avec une démarche différente, par Paul Dicks. Elle cherche à repérer un certain nombre de difficultés, de manques ou de privations dans différents domaines des conditions d'existence des ménages. Ces domaines peuvent renvoyer à une pauvreté de nature « existentielle » (alimentation, logement par exemple) ou de nature « sociale » (relations, emploi, loisirs,..). Certains auteurs ont proposé de recourir à différents critères, par exemple : un critère objectif, correspondant au fait que les pratiques examinées sont diffusées dans la majorité de la population et un critère social, correspondant au fait que les manques sont considérés comme défavorables ou inacceptables par la majorité de la population. Dans le domaine du logement, il peut s'agir du fait de ne pas avoir l'usage d'un WC ou d'une douche, ou habiter dans un logement humide. Pour l'alimentation, ne pas avoir mangé de viande ou de poisson la plupart des jours de la semaine. Dans le champ des privations sociales, avoir été au chômage, occuper un emploi précaire, ne pas pouvoir offrir des cadeaux...

La mise en œuvre de cette approche soulève de redoutables difficultés et réserve une large place au choix, nécessairement conventionnel, des chercheurs et des concepteurs d'enquête.

2.2.5. L'approche subjective

L'approche subjective consiste non pas à se référer à un seuil minimal de ressources défini conventionnellement ou à des conditions objectives d'existence, mais à interroger directement les ménages sur la perception qu'ils ont de ces réalités, à partir de question sur leur revenu, le revenu minimum nécessaire selon eux « pour joindre les deux bouts » et /ou leur degré « d'aisance financière ».

Différentes méthodes permettent ensuite, sur la base de ces réponses, d'établir un seuil de pauvreté subjectif. Les ménages dont le revenu est inférieur à ce seuil seront considérés comme pauvres (insécurité d'existence objective). Un autre indicateur (Insécurité d'existence subjective) consiste à dénombrer les ménages qui déclarent éprouver les difficultés financières à boucler leur budget.

Les réponses à ce type de questions ne sont pas dénuées d'ambiguïté; d'abord parce que, dans leurs réponses sur le revenu minimal nécessaire, la référence retenue par les ménages n'est pas nécessairement la même et peut traduire une aspiration plus qu'une nécessité. Ensuite, parce que la perception que les ménages ont de ces réalités peut être influencée par le besoin psychologique de minimiser les écarts entre les aspirations et la réalité. Les ménages les plus démunis, qui doivent faire avec peu de ressources, peuvent avoir ainsi tendance à minimiser les difficultés objectives qu'ils rencontrent dans leur vie quotidienne, ce qui risque de biaiser les résultats.

Dans les études empiriques concernant les différentes mesures de la pauvreté, plusieurs auteurs dans leurs analyses ont montré une forte corrélation entre les différentes approches de mesure.⁹ La présence de l'une des formes de pauvreté accroît sensiblement la probabilité d'occurrence de toutes les autres. Ceci nous paraît judicieux dans la mesure où, si l'on dispose par exemple d'un faible revenu alors, nos conditions de vie ne sauraient être favorables.

Dans la base de l'EESI, tout individu dont la modalité pour la Variable SIIND (Secteur Institutionnel) était 3 ou 4 était un acteur du secteur informel.

⁹ Les multiples facettes de la pauvreté dans un pays en développement : le cas de la capitale malgache : DIAL ; Document de Travail : DT/2001/07

CHAPITRE 3 : DEMARCHE METHODOLOGIQUE ET PRESENTATION DU MODELE DE REGRESSION UTILISE

En relation avec les objectifs de notre étude, ce chapitre présente la démarche méthodologique d'identification des déterminants de la pauvreté (c'est-à-dire ce qui caractérise la pauvreté) chez les acteurs du secteur informel et le modèle de régression utilisé pour modéliser la pauvreté. Il s'articule autour des étapes suivantes :

- construction des scores privatifs et choix d'un seuil de pauvreté ;
- identification des acteurs informels pauvres ;
- construction d'un modèle d'analyse.

3.1. CHOIX D'UN SEUIL DE PAUVRETE

Comme nous l'avons vu au chapitre deux, la pauvreté a plusieurs dimensions. Pour notre étude, compte tenu des données de l'enquête, il nous faut signaler que ces données ne nous permettaient pas de dire a priori qu'un individu était pauvre ou non ainsi, nous avons opté pour l'approche de mesure suivant les conditions de vie. C'est une approche qui consiste à repérer un certain nombre de manques chez un individu. Dans cette partie, nous construisons dans un premier temps des scores privatifs dans le but de faire un choix du seuil de pauvreté ; ensuite, nous sélectionnons à la base de la revue de la littérature et du bon sens, les variables qui expliquent le niveau de vie des populations. Enfin, nous procéderons à une modélisation à l'aide de la régression logistique.

3.2. CONSTRUCTION DES SCORES PRIVATIFS

L'approche par les « conditions d'existence » consiste à définir le seuil de pauvreté à partir des conditions matérielles d'habitation. Comme l'indicateur porte sur les stocks, il permet d'appréhender la pauvreté dans la durée, par opposition aux variables monétaires soumises à des variations conjoncturelles (comme le revenu ou les dépenses). Il s'agit d'un indicateur objectif non monétaire abordant la pauvreté sous l'angle des résultats plutôt que des moyens, même si le fait de disposer par

exemple d'un logement confortable peut aussi constituer un moyen d'améliorer par exemple sa productivité au travail et d'accroître son revenu.

3.2.1. Sélections des commodités

Les commodités retenues sont celles des variables sélectionnées parmi celles qui sont les plus caractéristiques des conditions et du cadre de vie des populations. Ces items présentent nécessairement une dépendance vis-à-vis du revenu. En effet, dans l'analyse, le manque d'une commodité doit être considéré comme une manifestation de la pauvreté.

Dix-huit commodités, parmi les plus révélatrices du niveau de vie, ont été retenues pour la construction du score de pauvreté d'existence. La liste complète de ces commodités est la suivante :

- matériau des murs ;
- matériaux du toit ;
- matériau du sol ;
- approvisionnement en eau de boisson ;
- type de lieu d'aisance ;
- éclairage ;
- énergie pour cuisine ;
- voiture ;
- téléviseur ;
- réfrigérateur ;
- réchaud à pétrole ;
- téléphone portable ;
- fer à repasser ;
- ordinateur ;
- réchaud à gaz ;
- poste radio ;
- ventilateur ;
- lecteur de DVD/CD vidéo.

La variable accès aux soins de santé qui est en fait une variable très caractéristique du niveau de vie manque dans les commodités retenues du fait de son absence dans la base de données.

3.2.2. Codage des variables qualitatives

Pour des besoins de simplifications de notre décompte des scores privatifs, nous avons procédé à un codage systématique des variables qualitatives que nous avons sélectionnées parmi les dix-huit commodités. Les valeurs accordées aux modalités de chaque variable sont consignées dans le tableau 4 suivant :

Tableau 4: Codage des variables qualitatives prise en compte dans l'étude

Variables	Valeur du codage	
	1	0
Matériau de murs	<ul style="list-style-type: none"> - Carabot - Brique simple de terre - Pisé, terre battue - Nattes, chaume, feuilles 	<ul style="list-style-type: none"> - Béton, parpaing, brique en terre cuite - Pierre de taille - Planche
Matériaux du toit	<ul style="list-style-type: none"> - Nattes - Chaume - Feuilles - Pailles 	<ul style="list-style-type: none"> - Ciment - Tôle - Tuile
Matériau du sol	<ul style="list-style-type: none"> - Bois - Terre 	<ul style="list-style-type: none"> - Ciment - Béton - carreaux
Approvisionnement en eau de boisson	<ul style="list-style-type: none"> - Puits - Rivière - Lac - Marigot - Bonne Fontaine gratuite - Revendeur d'eau de Robinet - Source non aménagée 	<ul style="list-style-type: none"> - Robinet individuel (abonnement individuel SNEC) - Robinet commun SNEC - Forage
Type de lieu d'aisance	<ul style="list-style-type: none"> - Pas de WC - Latrines non aménagées 	<ul style="list-style-type: none"> - Latrines aménagées - WC avec chasse eau
Eclairage	<ul style="list-style-type: none"> - Pétrole 	<ul style="list-style-type: none"> - Électricité AES SONEL - Groupe électrogène - Gaz
Energie pour cuisine	<ul style="list-style-type: none"> - Bois ramassé ou reçu - Sciure - Copeau de bois 	<ul style="list-style-type: none"> - Gaz - Electricité - Pétrole

3.2.3. Définition des scores privatifs

Un score privatif est une valeur numérique qui équivaut au nombre de manque (parmi les commodités retenues) qui est observé chez un individu.

Pour obtenir les scores privatifs, l'on commence par définir des « commodités de privations » à partir des commodités élémentaires retenus. Ainsi, chaque commodité élémentaire donne naissance à une « commodité de privation » qui vaut 1 si le manque est observé et 0 sinon. Le score privatif ou score de pauvreté en termes de conditions de vie est obtenu par sommation simple des « items de privation ». Cela suppose que l'on donne le même poids (on appelle famille de poids une famille sous la forme $\{P_i, P_i > 0 \text{ pour tout } i \text{ élément de } I \text{ où } I \text{ est un ensemble fini ou non et tel que: } \sum_{i \in I} P_i = 1\}$) aux différents manques subits par les individus. Il est cependant possible d'affecter aux items de privation des pondérations issues des données, mais nous n'adoptons pas ici cette démarche, étant donné que la pauvreté des conditions de vie concerne des scores suffisamment élevés pour éliminer le risque de mauvais classement. Un score privatif élevé traduit une précarité de conditions de vie et inversement. Un score privatif de 18 traduira une situation de pauvreté extrême des conditions de vie.

3.2.4. Seuil de pauvreté de qualité de vie.

L'étude de la pauvreté des conditions de vie ne peut conduire à un décompte des pauvres mais seulement à un examen des caractéristiques de ceux qui vivent dans l'inconfort. Dans cette étude, nous fixons de façon arbitraire un seuil de pauvreté des conditions de vie de manière à isoler une incidence de la pauvreté du même ordre que la pauvreté monétaire [11, 12, 13]. La distribution des scores de pauvreté d'existence est donnée par le tableau 5.

Tableau 5 : Répartition des individus par score de pauvreté de conditions de vie

Valeur du score	Pourcentages des individus	Pourcentages cumulés
17	9,4	9,4
16	14,2	23,6
15	11,3	34,9
14	9,7	44,6
13	7,2	51,8
12	5,9	57,7
11	5,1	62,8
10	5,7	68,5
9	6,5	75,0
8	5,3	80,3
7	5,3	85,6
6	5,0	90,6
5	4,1	94,7
4	3,0	97,7
3	1,6	99,3
2	0,6	99,9
1	0,1	100,0

La population étudiée a une taille de 38 599 sur lequel 234 variables ont été étudiées. Deux valeurs du tableau 5 (troisième colonne) encadrent le taux de pauvreté monétaire au niveau des individus qui était de 40,2% en 2001¹⁰. Nous optons pour la valeur supérieure¹¹ (44,6%), bien qu'elle soit un peu éloignée du taux de pauvreté monétaire. Ainsi, un individu sera considéré comme pauvre s'il ne dispose pas d'au moins 14 des 18 commodités retenues. Ainsi, la variable indicatrice de pauvreté que nous nommons « niveau de vie » vaut 1 si le score atteint 14 et 0 dans le cas contraire.

¹⁰ DSCN (2001), Résultats de l'ECAM II

¹¹ Il nous paraît en effet plus excusable de surestimer la pauvreté que de la sous-estimer du fait des implications que cela peut avoir sur les politiques de lutte contre la pauvreté.

3.2.5. Quelles sont les conditions de vie par secteur d'activité ?

Lorsqu'on s'intéresse au taux de pauvreté selon le secteur institutionnel, on observe que la pauvreté en termes des conditions de vie épargne relativement les deux secteurs d'activités formelles (cf. Tableau 6). Seulement 8,0 et 6,0 % d'individus vivent dans des conditions précaires respectivement dans le secteur public et dans le secteur privé formel. A l'opposé, le secteur agricole informel, avec 74,3 % d'individus vivant dans l'inconfort, est le secteur d'activité où les individus vivent le plus dans l'inconfort. Il est suivi du secteur informel non agricole. Par ailleurs, en associant les deux secteurs d'activité informels et en y faisant une évaluation de la pauvreté, on constate que 55,5% d'individus dont le secteur d'activité est le secteur informel agricole et non agricole sont pauvres.

Tableau 6 : Répartition des individus par secteur d'activité

Secteur institutionnel	Pourcentages des individus pauvres	Pourcentages des individus non pauvres
Public	8,0	92,0
Prive formel	6,0	94,0
Informel non agricole	25,1	74,9
Informel agricole	74,3	25,7

3.3. METHODES STATISTIQUES

Rappelons que dans notre travail, l'un des objectifs est de déceler les facteurs à l'origine de la pauvreté chez les acteurs du secteur informel. Comme les facteurs à l'origine de la pauvreté ne sont pas les mêmes chez les individus, il sera donc question de retenir de manière générale les facteurs qui font preuve d'un risque élevé. Dans cette section, nous décrivons les outils statistiques utilisés pour étudier ces facteurs.

Soit Y , le caractère binaire signalant pour un opérateur du secteur informel donné, son appartenance au groupe 0 ou 1.

A chaque i , on associe la variable aléatoire Y_i qui représente la valeur du caractère Y pour l'opérateur i .

$$Y_i = 1_{\text{groupe0}}(i) \quad \text{où } 1_{\text{groupe0}} \text{ est la fonction indicatrice du groupe0.}$$

Dans la suite, les acteurs informels appartenant au groupe0 ($Y=0$) sont non pauvres tandis que ceux du groupe1 ($Y=1$) sont pauvres.

Nous voulons mettre en évidence des variables X_1, \dots, X_p telle qu'on puisse affirmer de manière crédible que leur valeur affectent voire influencent celle de Y de façon significative parmi un certain nombre de variables que nous avons a priori choisi d'inclure dans notre étude. Ces variables ont été choisies sur la base de critères issus de la revue de la littérature mais aussi accessible dans la base de données disponible à l'INS. Une fois les variables X_1, \dots, X_p choisies, le problème revient à déterminer la loi de Y conditionnellement à X , où $X = (X_1, \dots, X_p)$.

Déterminer cette loi revient à estimer les paramètres $\pi(X)$ d'une loi de Bernoulli. Ayant observé $X=x$ sur un opérateur donné, on pourra dire que $\pi(x)$ représente le risque a priori que celui-ci se révèle être pauvre.

A chaque individu numéro i ($i=1, \dots, n$) des acteurs du secteur informel, on associe la variable aléatoire Y_i qui est une copie du caractère Y sur l'individu numéro i de la population ($Y_i \in \{0,1\}$).

$$Y_i = y_i, \text{ valeur du caractère } Y \text{ sur l'acteur numéro } i$$

Pour chaque $j=1, \dots, p$, on observe $X_{ji} = x_{ji}$ qui est donc la valeur de la variable X_j sur l'acteur numéro i .

C'est à partir de ces observations que nous allons estimer la loi de Y conditionnellement à X . Pour arriver à cette fin, nous allons développer dans les paragraphes suivants, l'approche logistique à travers l'usage du modèle logit ou logistique.

3.3.1. Choix des variables d'étude

Notre but étant d'obtenir des résultats aussi fiables et crédibles que possible, sur la base de critères socio-économiques disponibles dans la revue de littérature et la base de données de l'INS, nous avons porté notre attention sur les variables suivantes :

Sexe : nous voulons savoir s'il existe un lien entre le sexe et le niveau de vie. Nous voulons par exemple savoir si les femmes qui exercent dans le secteur informel sont plus pauvres que les hommes.

Age : nous voulons savoir si les jeunes sont plus vulnérables à la pauvreté que les vieux. Cette variable a été recodée en classe : [10 ; 24[, [24 ; 58[et plus de 58 ans.

Milieu de résidence : nous voulons savoir si en milieu rural et semi urbain, les actifs informels sont plus exposés à la pauvreté qu'en milieu urbain.

Niveau d'instruction : en introduisant cette variable dans notre étude, nous supposons que le niveau d'instruction doit avoir une influence sur le niveau de vie des individus.

Taille du ménage : nous entendons ici par taille du ménage le nombre d'individus qui constituent un ménage. Au regard de la revue de littérature, nous voulons savoir s'il existe une forte association entre la taille du ménage et le niveau de vie. Nous avons recodé cette variable comme suit : [0 ; 3[, [3 ; 9[et plus de 9 individus.

Type de ménage : comme pour la taille du ménage, on se propose d'étudier la corrélation entre le type de ménage et le niveau de vie.

Maison en location : pour la plupart d'individus, la possession des biens tels que des maisons en location crée des revenus supplémentaires. Ce qui peut expliquer de manière importante leur niveau de vie.

Catégorie socio-professionnelle : bien qu'a priori ceci peut prêter à confusion, la statistique descriptive de la base de données que nous disposons nous a montré que le secteur informel au Cameroun était bien organisé en catégorie socio professionnelle. Le tableau 7 nous donne une illustration des résultats. En introduisant cette variable dans notre étude, nous voulons par exemple vérifier l'information suivant laquelle les cadres sont moins vulnérables à la pauvreté par rapport aux ouvriers.

Type de ménage : comme pour la taille du ménage, on se propose d'étudier la corrélation entre le type de ménage et le niveau de vie

Nombre de personnes travaillant dans l'établissement : plus un établissement est rentable, plus il recrute d'employés et ils sont en général bien payés, ce qui améliore leur niveau de vie.

Le lieu où est exercé l'emploi secondaire : le lieu où s'exerce un emploi dépend du type d'emploi, voire la rémunération qui peut en découler.

Régime fiscal : le régime fiscal auquel est assujéti un établissement peut influencer le niveau de vie de ces employés. En effet, un établissement ne payant pas d'impôts ne connaît pas les mêmes problèmes qu'un autre qui les paye.

Tableau 7: Catégorie socio professionnelle selon le secteur institutionnel

		Secteur institutionnel					Total
		Public	Privé formel	Informel non agricole	Informel agricole	Inactif	
catégorie socioprofessionnelle	Cadre supérieur, ingénieur et assimilé	9,3%	3,0%	12,6%	12,8%	62,3%	100,0%
	Cadre moyen, agent de maîtrise	10,8%	7,4%	17,8%	16,6%	47,5%	100,0%
	Employé/ouvrier qualifié	2,4%	7,4%	33,4%	21,4%	35,5%	100,0%
	Employé/ouvrier semi qualifié	1,3%	5,9%	37,6%	22,4%	32,8%	100,0%
	Manoeuvre	,7%	1,4%	31,7%	25,2%	41,0%	100,0%
	Employeur	2,7%	8,9%	51,1%	16,2%	21,0%	100,0%
	Travailleur pour compte propre	1,3%	2,8%	34,5%	40,8%	20,6%	100,0%
	Aide-familial			13,7%	15,5%	70,8%	100,0%
	Apprenti		,5%	25,9%	37,9%	35,7%	100,0%
	Inclassable (décrivez)		22,1%	9,1%	31,0%	37,8%	100,0%
	Donnée manquante	2,9%		31,6%	19,9%	45,6%	100,0%
Total		3,4%	4,6%	31,3%	28,2%	32,5%	100,0%

3.3.2. Outils théoriques du modèle logit

Dans tout ce qui suit,

Y appartient à $\{0,1\}$ désigne une variable aléatoire binaire

X est une variable aléatoire à valeur dans un sous-espace E de IR^p .

Y sous la condition $X=x=(x_1, \dots, x_p)$ est une variable de Bernoulli $B(\pi(x))$ avec

$$\pi(x) = P(Y=1/X=x).$$

Y est une variable aléatoire représentant la survenue ou non d'un évènement codé en 0 et 1. Dans toute la suite, $Y=1$ matérialise le fait que l'on ait observé la pauvreté chez un individu $Y=0$ matérialise le contraire.

L'intérêt de ce chapitre est de spécifier $\pi(x)$ pour x fixé.

3.3.2.1. Rappel sur le modèle de régression linéaire

On dispose des observations $\{(x_i, z_i), i=1, \dots, n\}$ où $x_i \in \mathbb{R}^p$ est une variable explicative, et z_i est la variable à expliquer, $z_i \in \mathbb{R}$. Le modèle de régression linéaire se traduit par la relation perturbée suivante :

$$z_i = x_i \beta + \varepsilon_i, \quad i = 1, \dots, n,$$

où $\beta = (\beta_1, \dots, \beta_p)$ est la transposée du vecteur ligne $(\beta_1, \dots, \beta_p)$, et où les hypothèses suivantes sont faites :

- $H_1 : E(\varepsilon_i / x_i) = 0$ (espérance de ε_i conditionnellement à x_i) ;
- $H_2 : Var(\varepsilon_i) = \sigma^2$ (variance de ε_i) ;
- $H_3 : Cov(\varepsilon_i, \varepsilon_j) = 0$ pour $i \neq j$

avec ε_i , pour i fixé, un bruit blanc (BB) encore appelé résidu. Les paramètres du modèle sont β et σ^2 . Ajuster le modèle revient donc à donner une estimation de β et σ^2 .

Limite du modèle linéaire standard pour notre étude

Supposons que la variable dépendante Z est une variable continue à valeurs dans $[0,1]$. La régression linéaire peut s'étendre à l'infini lorsque la valeur de la variable indépendante s'accroît jusqu'à l'infini, alors qu'une probabilité, par définition doit se situer entre 0 et 1. Ainsi, on a recours au modèle Logit qui, est un cas particulier des modèles linéaires généralisés.

3.3.2.2. Nécessité d'un autre modèle de régression : le modèle logistique

Ici, on cherche à estimer $P(Y=1/X=x) = \pi(x)$. Dans la suite, nous noterons parfois π au lieu de $\pi(x)$, qui est donc la probabilité pour un individu d'être pauvre, et on a alors : $0 \leq \pi \leq 1$. Le problème posé est qu'une probabilité est bornée à droite et à gauche. Dans le modèle logistique, on essaie d'abord de se libérer de cette entrave. On procède alors de la façon suivante :

- On divise π par $1-\pi$, ainsi la borne à droite est supprimée, puisque, quand π tend vers 1^- $\frac{\pi}{1-\pi}$ tend manifestement vers $+\infty$. si π tend vers 0, alors $\frac{\pi}{1-\pi}$ tend aussi vers 0.

- On applique alors une seconde transformation, du type logarithmique cette fois de telle sorte que pour π tendant vers 0^+ , la transformation tend vers $-\infty$. Cette double transformation est appelée, en terme technique, Logit de π .

$$\text{Logit}(\pi) = \ln\left(\frac{\pi}{1-\pi}\right) \text{ et est à valeurs dans } \mathbb{R} \cup \{-\infty, +\infty\}.$$

La fonction $K(x) = \text{Logit}(\pi)$ est appelée une *link function* (*fonction de lien en français*) dans la théorie des modèles linéaires généralisés¹².

Remarquons que la transformation inverse de $\text{Logit}(\pi)$ est donnée par :

$$\forall s \in \mathbb{R}, \quad G(s) = \frac{\exp(s)}{1 + \exp(s)},$$

qui n'est rien d'autre que la fonction de répartition de la loi logistique, d'où la terminologie modèle logistique. C'est une fonction dont le graphe sur \mathbb{R} admet le point $(0, \frac{1}{2})$ pour centre de symétrie.

3.3.2.3. Définition du modèle Logit

Le modèle Logit est défini par :

$$\pi(x) = P(Y = 1 / X = x) = G(\beta x) \text{ où } \text{Logit}(\pi(x)) = \beta x.$$

3.3.2.4. Importance du modèle Logit : l'Odds Ratio

Si le modèle logit est fréquemment utilisé, c'est à cause de l'interprétation de l'exponentielle du coefficient d'une co-variable comme un « Odds Ratio » (Rapport de chance en français).

On entend par « Odds Ratio » le quotient

$$OR(x_i) = \frac{\pi(x_i)}{1 - \pi(x_i)} = \frac{p(Y_i = 1/x_i)}{p(Y_i = 0/x_i)} = e^{x_i \beta},$$

$OR(x_i)$ est le rapport de chances sous x_i .

Le modèle logit traduit que le log du rapport des chances suit le modèle linéaire $x_i \beta$. L'interprétation de β est la suivante :

¹² Mc Cullagh et Nelder (1989)

Pour une variable qualitative, il y a $e^{x_i\beta}$ fois plus de chances que l'événement $y_i = 1$ se réalise qu'il ne se réalise pas.

Soit x une variable binaire, par exemple la variable sexe, alors l'odds ratio se définit comme suit :

$$OR(sexe) = \frac{\frac{p_1}{1-p_1}}{\frac{p_0}{1-p_0}},$$

où p_0 représente la probabilité d'être pauvre pour une femme et p_1 la probabilité d'être pauvre pour un homme. Ainsi,

Si $OR=1$, alors le risque d'être pauvre n'est pas lié au sexe.

Si $OR>1$, cela signifie que le risque d'être pauvre est plus élevé chez les hommes que chez les femmes.

C'est le cas contraire si $OR<1$.

3.3.2.5. Estimation du modèle

Elle se fait par la méthode du maximum de vraisemblance que nous décrivons ci-dessous. On commence toujours par écrire la vraisemblance de l'échantillon observé de la variable dépendante conditionnement aux valeurs observées des variables explicatives. En cas de données individuelles $\{(y_i, x_i), i=1, \dots, n\}$, cette vraisemblance vaut :

$$Ln(\beta) = \prod_{i=1}^{i=n} \pi_i^{y_i} (1 - \pi_i) = \prod_{i=1}^{i=n} (1 - \pi_i) \exp\{y_i \times \text{Logit}(\pi_i)\},$$

où $\pi_i = G(\beta x_i)$ est associée à l'individu numéro i .

On a utilisé le fait que si Z est une variable aléatoire qui suit une loi de Bernoulli de paramètre p alors :

$$P(Z = z) = p^z (1 - p)^{1-z} \text{ où, bien entendu, } z \in \{0,1\}.$$

Ecriture de la log-vraisemblance

La log-vraisemblance $\ln(\beta) = \log Ln(\beta)$ vaut alors :

$$\ln(\beta) = \langle T_x(y), \beta \rangle + \Psi_x(\beta),$$

avec :

$$T_x(y) = \sum_{i=1}^n y_i x_i \text{ et } \Psi_x(\beta) = - \sum_{i=1}^n \log(1 - \exp \langle \beta, x_i \rangle).$$

Dans ces expressions, \langle, \rangle désigne le produit scalaire sur IR^p :

si $a \in IR^p$ et $b \in IR^p$, $\langle a, b \rangle =^t ab$.

Contrairement au modèle de régression linéaire où l'on sait déterminer une estimation de β de manière explicite, ici l'on détermine une estimation β à base d'un algorithme numérique. On note $\hat{\beta}_n$, l'estimateur du maximum de vraisemblance de β .

Proposition

1. La fonction $\beta \rightarrow \ln(\beta)$ est strictement concave
2. $I_n(\beta) = \sum x_i^t x_i (1 - \pi_i)$, où on a toujours $\pi_i = \pi(x_i) = G(\exp^t x_i \beta)$.
3. la loi asymptotique de :

$\hat{\beta}_n - \beta$ est $N_p(0, (I_n(\beta))^{-1})$, β identifiable, c'est - à - dire que si $\beta \neq \beta'$ les lois de $\{(Y_i / x_i), i = 1, \dots, n\}$ associées à β et β' sont différentes.

Ici, I_n désigne la matrice d'information de Fisher.

Rappel sur la méthode du maximum de vraisemblance

Soit X , une variable aléatoire obéissant à une loi de probabilité de paramètre $\theta: f(X, \theta)$. Ayant observé $X_1 = x_1, \dots, X_p = x_p$, on voudrait estimer la valeur inconnue du paramètre θ à partir de ces observations.

Principe de la méthode : on estime cette valeur par celle qui maximise la probabilité d'obtenir les valeurs observées sur X .

Cette méthode suppose :

La définition de $FV(\theta)$, dite fonction de vraisemblance, et qui permet de décrire la probabilité d'obtenir les valeurs observées :

$$FV(\theta) = f(X_1, \theta) \times f(X_2, \theta) \times \dots \times f(X_n, \theta) = \prod_{i=1}^n f(X_i, \theta).$$

Dans le cas d'une variable discrète qui va nous concerner dans notre travail, $FV(\theta)$ est un produit de probabilités. La fonction $FV(\theta)$ atteint son maximum en la valeur de θ qui annule sa dérivée première :

$$\frac{\partial FV(\theta)}{\partial \theta} = 0.$$

En pratique, afin de rendre la tâche moins difficile, on opère sur la transformation logarithmique. On pose alors :

$$L(\theta) = \log FV(\theta).$$

$L(\theta)$ est appelé *log-vraisemblance*. Ainsi, l'estimation du maximum de vraisemblance de θ est aussi la valeur θ qui annule la dérivée première de $L(\theta)$:

$$\frac{\partial L(\theta)}{\partial \theta} = 0.$$

3.3.3. Evaluation et validation du modèle

La qualité d'un modèle statistique se conçoit en général sur la significativité individuelle des coefficients, sur la significativité globale du modèle au moyen des tests de vraisemblance, l'analyse des résidus et sur la puissance de discrimination de ce modèle.

3.3.3.1. Test des paramètres du modèle

Avant toute régression, il est indispensable de s'assurer que les variables explicatives introduites dans le modèle sont indépendantes entre elles. Le test de Khi-deux est un test d'indépendance bien approprié à cet effet. Le logiciel STATA, que nous avons utilisé pour notre étude statistique du modèle envoie un message d'erreur en cas de dépendance de deux variables explicatives après la régression.

Pour tester la significativité d'un ou plusieurs coefficients du modèle, on utilise soit la statistique de Wald W , soit le test de Rapport de Vraisemblance LR.

3.3.3.2. Test de Wald

Pour un seul coefficient, la statistique de Wald s'écrit :

$$W = \frac{\hat{\beta}_k}{\widehat{es}(\hat{\beta}_k)} \xrightarrow{\beta_k \rightarrow 0} N(0,1).$$

Pour plusieurs coefficients, par exemple :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_q = 0.$$

La statistique de Wald s'écrit :

$$W = \hat{\beta}' (\widehat{V}(\hat{\beta}))^{-1} \hat{\beta} \xrightarrow{Loi} \chi^2(q).$$

3.3.3.3. Test de Rapport de Vraisemblance

La statistique du RV pour un seul coefficient, s'écrit :

$$2 LR = -2 \text{Log}(L_c/L) \xrightarrow{Loi} \chi^2(1)$$

Pour plusieurs coefficients, par exemple :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_q = 0,$$

la statistique du RV s'écrit :

$$LR = -2\text{Log}(Lc/Lc^-) \xrightarrow{\text{Loi}} \chi^2(q) \text{ où } Lc \text{ est la vraisemblance évaluée sous la}$$

contrainte H_0 et Lc^- la vraisemblance sans contrainte.

3.3.3.4. L'adéquation du modèle aux données « Goodness of fit »

Une fois que les étapes du choix des co-variables et de leurs formes fonctionnelles ont été effectuées, on peut déterminer la qualité de l'ajustement du modèle aux données ou, en anglais, le « Goodness of fit ».

Pour fixer les idées, notons les valeurs observées de l'outcome $y'=(y_1, \dots, y_n)$ et les valeurs prédites par le modèle $\hat{y}'=(\hat{y}_1, \dots, \hat{y}_n)$, où n est la taille de l'échantillon. On considérera que l'ajustement est satisfaisant si :

1. La distance entre l'outcome observé y et l'outcome prédit par le modèle \hat{y} est petite.
2. Le modèle est bien « calibré », c'est-à-dire les fréquences prédites sont proches de celles observées.
3. Le modèle permet de bien discriminer entre les opérateurs non pauvres et ceux qui sont pauvres en fonction des variables explicatives c'est-à-dire on obtient de bonnes sensibilités et spécificités.

Pour nous aider dans cette tâche, nous allons nous appuyer sur des tests de « Goodness of fit » comme, le test de Hosmer et Lemeshow, la statistique de Pearson, sur l'analyse des résidus comme, les résidus de Pearson, ainsi que sur l'évaluation de la capacité à discriminer les outcomes positifs $y = 1$ des outcomes négatifs $y = 0$ par l'inspection de la courbe ROC.

La démarche que nous allons adopter consiste à évaluer, d'abord, globalement l'adéquation du modèle au moyen des différents tests de « Goodness of fit », puis, en principe lorsqu'on est satisfait de la qualité de l'ajustement global, à déterminer s'il n'y a pas localement des observations très mal ajustées et ayant possiblement un effet important sur l'estimation des coefficients. Le but des ces évaluations globale et locale est de s'assurer que l'ajustement du modèle soit satisfaisant pour toutes les valeurs observées dans l'échantillon des variables explicatives. Finalement, l'évaluation du

pouvoir discriminant du modèle nous permettra d'appréhender si nous avons choisi les « bonnes » variables explicatives ou s'il manque d'importants régresseurs pour arriver à prédire avec suffisamment de précision l'outcome.

3.3.3.5. Evaluation de la calibration du modèle : le test de Hosmer et Lemeshow

Le test de Hosmer et Lemeshow est basé sur un regroupement des probabilités prédites par le modèle, par exemple par décile. On calcule, ensuite, pour chacun des groupes le nombre observé de pauvres et de non pauvres, que l'on compare au nombre espéré prédit par le modèle. On calcule alors une distance entre les fréquences observées et prédites au moyen d'une statistique du Khi-deux. Lorsque cette distance est petite on considère que le modèle est bien calibré.

Ce test comporte les étapes suivantes :

- Estimer le modèle ;
- Obtenir les $\hat{\pi}_i = Pr(Y_i = 1/X)$;
- Ordonner dans le sens croissant les $\hat{\pi}_i$;
- Partager les $\hat{\pi}_i$ en G groupes (quintile, décile ou percentile) ;
- Calcul de la statistique.

n_g = nombre d'observations dans le groupe g , $g = 1, 2, \dots, G$

$$\bar{y}_g = \frac{\sum_{i=1}^G y_i}{n_g}, \quad \bar{\pi}_g = \frac{\sum_{i=1}^G \pi_i}{n_g}.$$

$$HL = \sum_{i=1}^n \frac{(n_g \bar{y}_g - n_g \bar{\pi}_g)^2}{n_g \bar{\pi}_g (1 - \bar{\pi}_g)} \sim \chi^2(G - 2).$$

- Calcul de la p -value : $p\text{-value} = Pr(HL > \chi^2(G-2))$.

Si $p\text{-value} < \alpha = 0.05$, alors le modèle n'est pas adéquat aux données ou bien la qualité de l'ajustement est douteuse.

NB : les statistiques de Hosmer et Lemeshow (HL) dépendent du nombre de groupes choisis. C'est pour cela qu'eux-mêmes le recommandent comme un indicateur et non comme un test proprement dit. C'est pour cela qu'avant de prendre la décision que le

modèle est adéquat, ou non, il faut lui associer les critères tels que AKAIKE (AIC) et BIC, et faire la courbe de lissage.

3.3.4.6. L'analyse des résidus

Un résidu est une mesure de la distance entre l'outcome observé y et l'outcome prédit par le modèle \hat{y} . Comme on va le voir, il existe plusieurs définitions de résidus et chacune d'elle correspond à un concept particulier de distance.

Le but de l'analyse des résidus est multiple :

- 1) il s'agit de vérifier qu'il n'y a pas des erreurs systématiques ;
- 2) de déterminer s'il y a des observations très mal expliquées (résidus extrêmes) ;
- 3) et si certaines observations ont un effet important de levier sur les résultats des estimations.

Comme chaque observation a son résidu associé il y a autant de résidus que d'observations.

L'on considérera donc des mesures globales résumant l'ensemble des résidus par un seul chiffre et permettant ainsi d'apprécier l'ajustement global du modèle aux données (autrement dit on résume la distance entre y et \hat{y}), ainsi que des mesures locales fournies par chacun des résidus et permettant de vérifier que la contribution à la mesure globale de chacune des observations est plus ou moins équivalente.

Eventuellement, certaines valeurs caractéristiques d'une variable explicative peuvent être éliminées de l'analyse si elles sont jugées trop « loin » du nuage de points et qu'elles ont un effet de levier important sur l'estimation des coefficients. On cherche, en définitive, à ajuster le modèle sur le centre de gravité du nuage de points et il n'est pas désirable que quelques valeurs extrêmes (qui peuvent être des erreurs de mesure ou des cas complètement atypiques) modifient sensiblement les estimations.

3.3.3.7. Le Résidu de Pearson est donné par la formule suivante :

$$r_i = \frac{y_i - \pi_i}{\sqrt{\pi_i(1 - \pi_i)}}, i=1, \dots, n$$

Ce résidu sera d'autant plus grand que le nombre de pauvres prédit est différent du nombre observé et que le dénominateur est petit. De plus son interprétation est délicate en ce sens qu'elle dépend de l'amplitude $y_i - \pi_i$ et de la variance $\pi_i(1 - \pi_i)$ de sorte que le résidu de Pearson n'est pas standardisé. Pour cela, nous définissons le résidu de Pearson standardisé.

Le résidu de Pearson standardisé est défini comme :

$$r_i^s = \frac{r_i}{\sqrt{1 - h_{ii}}}; \forall i = 1, \dots, n.$$

Avec

$$h_{ii} = \pi_i(1 - \pi_i)^t X_i \text{Var}(\beta) X_i.$$

3.3.4. Evaluation du pouvoir discriminant du modèle : sensibilité, spécificité et courbe ROC

Dans le but de construire le modèle de prédiction, il est d'un intérêt de faire un examen de son pouvoir discriminant. Cet examen se fait par calcul de l'aire au-dessous de la courbe ROC (Receiver Operating Characteristic) ou courbe de caractéristiques d'efficacité. Cette courbe permet d'étudier les variations de la spécificité et de la sensibilité d'un test pour différentes valeurs du seuil de discrimination. Sur l'axe des abscisses, on porte la variable 1- spécificité donnant l'effectif (en pourcentage) de non pauvres parmi les pauvres, sur l'axe des ordonnées, on place la sensibilité égale à l'effectif (en pourcentage) de vrais pauvres parmi les non pauvres. Si Se et Sp désignent respectivement la sensibilité et la spécificité du test, nous avons :

$Se = Pr(\text{le test décide que l'individu est pauvre sachant qu'il est effectivement pauvre})$;

$Sp = Pr(\text{le test décide que l'individu n'est pas pauvre sachant qu'il n'est effectivement pas pauvre})$.

La surface sous cette courbe nous permet d'évaluer la précision du modèle pour discriminer les opérateurs pauvres des opérateurs non pauvres.

On retiendra comme règle du pouce :

- Si aire ROC = 0.5 il n'y a pas de discrimination ;
- Si $0.5 \leq \text{aire ROC} < 0.7$ la discrimination est insuffisante ;
- Si $0.7 \leq \text{aire ROC} < 0.8$ la discrimination est acceptable ;
- Si $0.8 \leq \text{aire ROC} < 0.9$ la discrimination est excellente ;
- Si aire ROC ≥ 0.9 la discrimination est exceptionnelle.

CHAPITRE 4: PRÉSENTATION DES RESULTATS

Dans cette partie, après avoir présenté les résultats de notre analyse nous procédons ensuite à leurs interprétations. Tous les résultats présentés ici sont obtenus à partir des Logiciels STATA version 8.0 et SPSS version 12.0

Les variables d'intérêt ayant fait l'objet de la modélisation ont été choisies selon les critères définis au paragraphe intitulé « méthodes statistiques » et au chapitre Revue de la Littérature. Le test de Khi-deux a permis de tester la liaison entre chacune des variables explicatives et la variable niveau de vie. Ensuite nous avons testé l'indépendance des variables explicatives entre elles. Ceci nous a conduit à une réduction du nombre de variables et nous a permis d'éviter les risques de surestimation et de multicolinéarité (on peut dire qu'il y a multicolinéarité entre plusieurs variables si l'une des variables s'écrit comme combinaison linéaire des autres variables).

Nous postulons donc pour le modèle suivant :

$$\text{Logit}(\pi) = \beta_0 + \beta_1 \text{sex}e + \beta_2 \text{agee} + \beta_3 \text{g5} + \beta_4 \text{nivinst} + \beta_5 \text{taille} + \beta_6 \text{typem} + \beta_7 \text{e24} + \beta_8 \text{ap3} + \beta_9 \text{as8} + \beta_{10} \text{ap6a} + \beta_{11} \text{ap5} + \beta_{12} \text{as7a},$$

où *agee*, *g5*, *nivinst*, *taille*, *typem*, *e24*, *ap3*, *as8*, *ap6a* et *ap5* désignent respectivement le l'âge, le milieu de résidence, le niveau d'instruction, la taille du ménage, le type de ménage, le nombre de maisons en location, la catégorie socio professionnelle, type de local où est exercé l'emploi, régime fiscal de l'emploi principal, le nombre de personnes travaillant dans l'établissement, régime fiscal de l'emploi secondaire.

Le problème est de déterminer la combinaison des variables qui décrivent au mieux la probabilité d'être pauvre chez les actifs du secteur informel. Cela revient à estimer les paramètres β_i , $i=0, \dots, 12$.

4.1. ESTIMATION

La construction d'un modèle est d'une grande importance. En effet, à partir d'un modèle décrivant nos données, nous espérons pouvoir inférer des caractéristiques

de la population d'intérêt, comme par exemple : l'association entre l'âge et le niveau de vie.

Les variables candidates retenues ont été utilisées comme co-variables pour la régression logistique multivariable, avec comme variable dépendante « niveau de vie » (code : 1 si l'individu est pauvre et 0 sinon). Notre objectif étant d'identifier les co-variables qui sont associées au niveau de vie chez les acteurs du secteur informel.

4.2. RESULTAT DE L'ESTIMATION

Nous avons obtenu le modèle final après une évaluation pas à pas. Le modèle obtenu est le suivant :

$$\text{Logit}(\pi) = -1.54 - 0.17 \times \text{agee} - 0.312 \times \text{taille} - 0.38 \times \text{sexe} + 0.248 \times \text{ap3} + 1.23 \times \text{g5} - 1.11 \times \text{e24} - 0.89 \times \text{nivinst} + 0.03 \times \text{as8},$$

où *agee*, *taille*, *ap3*, *g5*, *e24*, *nivinst* et *as8* désignent respectivement l'âge, la taille du ménage, la catégorie socio professionnelle, le milieu de résidence, le nombre de maisons en location, le niveau d'instruction et le type de local où est exercé l'emploi.

Le test de Wald a ensuite été effectué et a montré que les coefficients retenus dans le modèle étaient significativement différents de zéro au seuil de 5%.

Tableau 8 : Estimation des paramètres du modèle

niveaude	Coef.	Std. Err.	z	P> z	IC 95 % ¹³
agee	-0.170	0.072	-2.33	0.020	[-0.313 ; -0.027]
taille	-0.312	0.067	-4.64	<10 ⁻³	[-0.443 ; -0.180]
sexe	-0.383	0.087	-4.38	<10 ⁻³	[-0.555 ; -0.212]
ap3	0.248	0.050	4.97	<10 ⁻³	[0.150 ; 0.346]
g5	1.232	0.055	22.18	<10 ⁻³	[1.123 ; 1.341]
e24	-1.111	0.188	-5.89	<10 ⁻³	[-1.481 ; -0.741]
nivinst	-0.895	0.058	-15.37	<10 ⁻³	[-1.01 ; -0.781]
as8	0.036	0.014	2.51	0.012	[0.008 ; 0.064]
cons	-1.541	0.460	-3.35	0.001	[-2.443 ; -0.639]

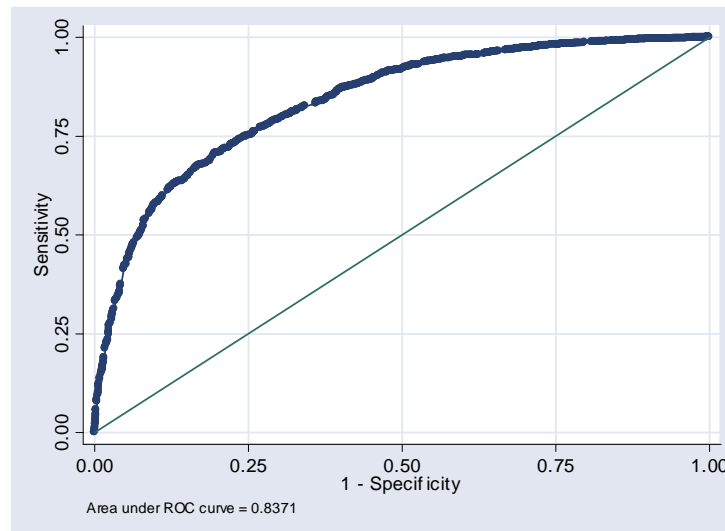
Ces résultats indiquent que les variables qui sont caractéristiques du niveau de vie sont : l'âge, la taille du ménage, le sexe, la catégorie socio-professionnelle, le milieu de résidence, le nombre de maison en location, le niveau d'instruction, le type de local où est exercé l'emploi.

4.2.1. Adéquation du modèle

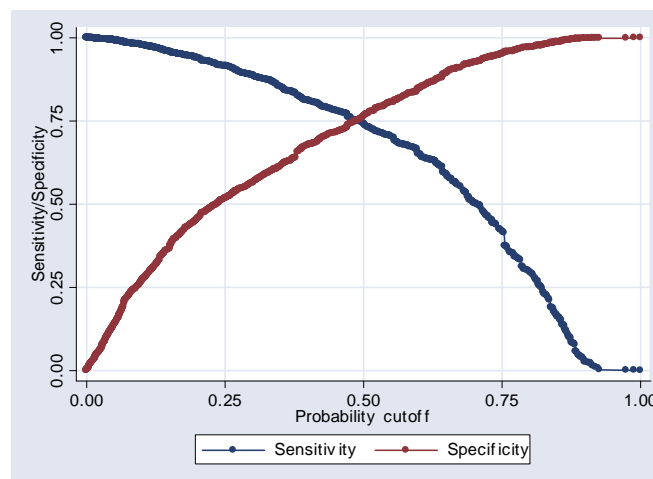
Dans le cas de ce travail, nous nous sommes limités à calculer la valeur de l'aire au dessous de la courbe ROC, car elle est nécessaire pour examiner le pouvoir discriminant du modèle.

Pour ce modèle, l'aire sur la courbe ROC vaut 83.71%, ce qui traduit un pouvoir discriminant excellent. La statistique R² de Mac Fadden vaut 0,4065 ce qui explique 40.65%, de déviance du modèle : ceci traduit une performance acceptable du modèle étant donné que nous sommes en présence des données d'enquête.

¹³ Intervalle de confiance à 95 %

Figure 2 : Courbe ROC

Le seuil de discrimination est de 0.49 (voir figure 3) propre aux données correspond au point où la sensibilité est égale à la spécificité contrairement à la valeur 0.5 généralement prise comme valeur par défaut. A ce seuil, la classification fournie par le modèle est bonne soit **75.91%**.

Figure 3 : Graphe des courbes de sensibilités et spécificités

CALIBRATION DU MODELE

Dans notre étude, nous avons regroupé les individus en 10 groupes et, le test de Hosmer et Lemeshow a donné une p-value de 0.561. On constate que dans le cas du

modèle retenu, ce test donne un résultat convaincant par conséquent, l'ajustement global du modèle aux données est satisfaisant.

4.2.3. Analyse des résidus

En plus de l'étude de l'adéquation, de la classification, et du calibrage du modèle qui fournissent de bonnes performances, une analyse des résidus de Pearson est effectuée afin de vérifier qu'il y a pas d'erreur systématique et de déterminer qu'il n'y a pas des observations très mal expliquées (résidus extrêmes). Rappelons qu'un résidu est une mesure de la distance entre la valeur observée y et la valeur prédite \hat{y} : il y a autant de résidus que d'observations. Ainsi, d'après la figure 4 qui ressort le nuage des résidus standardisés de Pearson en fonction des probabilités prédites par le modèle, nous ne constatons pas de grand déséquilibre dans la distribution des résidus, néanmoins, les individus numérotés: 1409, 14303 sont des observations mal expliquées (aberrantes c'est-à-dire que ces observations sont sources d'instabilité pour les paramètres estimés et ont été supprimées afin d'améliorer les performances du modèle), environ 95% des points sont dans l'intervalle $[-2 ; 2]$.

- Il est presque évident que la classe sociale d'un individu influence de façon très significative son niveau de vie. Ainsi, la catégorie socioprofessionnelle laisse présager une vulnérabilité à la pauvreté très élevée chez les manœuvres, les apprentis et chez les ouvriers non qualifiés par rapport aux cadres supérieurs.
- De la description de notre modèle, nous notons que les individus ayant un niveau d'instruction élevé (supérieur) présentent moins de risque d'être pauvre contrairement à ceux ayant un niveau d'instruction faible (primaire et moins).
- Il faut également noter que les acteurs du secteur informel vivant en milieu rural et semi urbain présentent respectivement 10 et 2 fois plus de risque d'être pauvre par rapport à ceux vivant en milieu urbain. Ceci n'est pas du tout surprenant quand on sait qu'au Cameroun comme dans beaucoup de pays d'Afrique subsaharienne, l'activité économique est consacrée dans la plus part des grandes métropoles qui sont en générale située en milieu urbain.
- Bien que nous pensions que le lieu où est exercé l'emploi pouvait avoir une influence sur le niveau de vie des acteurs du secteurs informels, il ressort de cette étude que cette variable ne contribue aucunement à l'explication du niveau de vie des acteurs du secteurs informel au Cameroun.

CHAPITRE 5 : CONCLUSION ET RECOMMANDATIONS

Notre travail consistait principalement à identifier les déterminants de la pauvreté chez les actifs du secteur informel ; tout en décrivant leur situation. L'étude que nous avons menée a porté sur la totalité des données issues de la phase 1 de l'enquête EESI.

Le problème fondamental était de trouver un modèle permettant d'expliquer le niveau de vie des actifs du secteur informel. Ce modèle devait intégrer les caractéristiques fondamentales du niveau de vie des acteurs du dit secteur. Il s'agit du modèle logistique qui prend en compte une variable dépendante catégorielle. Notre recherche s'est déroulée en deux étapes bien distinctes :

- la première étape était consacrée à l'identification des pauvres. Pour ce faire, nous avons construit des scores de privation ensuite nous avons choisi un seuil de pauvreté. Nous avons vu que le secteur informel avec 55,5% de pauvres, était le secteur institutionnel le plus vulnérable à la pauvreté des conditions d'existence au Cameroun ;
- la seconde étape consistait à identifier les facteurs de la pauvreté. Partant du modèle logistique, et à l'aide de la revue de la littérature, nous avons sélectionné les variables qui pourraient a priori justifier le niveau de vie d'un individu exerçant dans le secteur informel.

L'observation, même rapide des activités du secteur informel au Cameroun permet de constater qu'elles deviennent du jour au jour des activités plus complexes et de plus en plus variables. Cette complexité fait naître de multiples problèmes qu'il faut résoudre afin de permettre son développement et l'épanouissement des ses acteurs.

Parmi les variables sur lesquelles il faut agir, il y a la scolarisation. Il s'agit de développer les interventions visant à améliorer l'accès des actifs informels à l'éducation. En effet, les éléments d'analyse précédemment mis en évidence suggèrent un profond rehaussement très significatif de la scolarisation primaire et post-primaire, en particulier pour les jeunes ; ceci devrait leur permettre de pouvoir enrichir leur capital intellectuel humain.

Il faut également favoriser un développement des activités économiques en milieu rural et semi urbain; ceci dans le but de diminuer le risque de pauvreté dans ces milieux de résidence.

L'Etat devrait mettre la pression aux entreprises pour qu'elles puissent donner le minimum nécessaire aux ouvriers afin qu'ils puissent avoir des conditions de vie acceptables.

Cette étude bien qu'intéressante et profitable, ne devrait pas se limiter à ce niveau. Des études beaucoup plus sophistiquées doivent être envisagées compte tenu de l'ampleur de la situation. Ces études devront intégrer tous les paramètres nécessaires.

BIBLIOGRAPHIE

1. E. P. NDONG NGUEMA, (2006), Cours de Data Mining, ENSP, Université de Yaoundé I, Cameroun.
2. P. KOUEKAM, (2006), Mémoire de Master de Statistique Appliquée, Université de Yaoundé I, Cameroun
3. E. NDORUHIRWE, (2005) Mesure, analyse et caractérisation de la pauvreté infantile en Guinée, Mémoire de Maîtrise en Economie, Université Laval, Canada.
4. C. OLM et M. SIMON, (2005), Une approche de la pauvreté à partir de l'hétérogénéité des conditions de vie. CAHIER DE RECHERCHE N° 214, Paris France.
5. INS, (2005), EESI : Document de méthodologie, Cameroun.
6. INS, (2005), Rapport principal phase 1 de l'ESSI, Cameroun.
7. Patrick Taffé, (2004), Cours de Régression Logistique Appliquée. Institut Universitaire de Médecine Sociale et Préventive et Centre d'Epidémiologie Clinique. Lausanne, 64 P.
8. I. Aminot, MN. Damon, (2002), Régression logistique: intérêt dans l'analyse de données relatives aux pratiques médicales, Revue Médicale de l'Assurance Maladie volume 33 n°2.
9. DIAL, (2001), Les multiples facettes de la pauvreté dans un pays en développement : le cas de la capitale malgache, Document de Travail.
10. Rapport sur le développement dans le monde 2000/2001 « Combattre la pauvreté ».
11. P. Traissac, Y. Martin-Prével, F. Delpeuch et al., (1999), Régression logistique vs autres modèles linéaires généralisés pour l'estimation de rapports de prévalences, Revue d'Epidémiologie et de Santé publique, 47 :593-604.
12. S. LOLLIVIER et D. VERGER, (1997), Une approche de la pauvreté par les conditions de vie, Document de travail, N° 97-01, Direction des Statistiques Démographiques et Sociales de l'INSEE.

- 13.S. LOLLIVIER et D. VERGER, (1996), Les indicateurs de privation des ménages en 1994 dans le panel européen, Document de travail, N° 97-06, Direction des Statistiques Démographiques et Sociales de l'INSEE.
- 14.B. NOLAND et C. T. WHELAN, (1996), Resources, Definitions an the measurement of poverty, Oxford Clarendon Press.
- 15.DNSC, (1993), Le secteur informel à Yaoundé: Principaux résultats Enquête 1-2-3 1993 phase 2, Cameroun.
- 16.DW Hosmer and S.Lemeshow, (1989), Applied logistic regression. John Wiley & Son.
17. A. Albert and JA Anderson, (1984), On the existence of maximum likelihood estimates in logistic regression models, *Biometrika*. 71:1-10.