

DOCUMENT DE TRAVAIL

DT/2005-04

# Inégalités des chances sur le marché du travail : effets de l'origine sociale sur la mobilité occupationnelle à Lima

*Laure PASQUIER-DOUMER*  
*Gerardo David ROSAS SHADY*

DIAL • 4, rue d'Enghien • 75010 Paris • Téléphone (33) 01 53 24 14 50 • Fax (33) 01 53 24 14 51  
E-mail : [dial@dial.prd.fr](mailto:dial@dial.prd.fr) • Site : [www.dial.prd.fr](http://www.dial.prd.fr)

# INEGALITES DES CHANCES SUR LE MARCHE DU TRAVAIL : EFFETS DE L'ORIGINE SOCIALE SUR LA MOBILITE OCCUPATIONNELLE A LIMA<sup>1</sup>

Laure Pasquier-Doumer  
DIAL, IEP, Université de Rouen  
[pasquier@dial.prd.fr](mailto:pasquier@dial.prd.fr)

Gerardo David Rosas Shady  
DIAL, TEAM/CED, Université Paris I  
[davidro@consultant.iadb.org](mailto:davidro@consultant.iadb.org)

**Document de travail DIAL**  
Mars 2005

## RESUME

La société péruvienne, caractérisée par une répartition très inégalitaire des revenus, est-elle également une société injuste au regard du critère d'égalité des chances ? L'article cherche des éléments de réponse à cette question en s'interrogeant sur le rôle de l'origine sociale sur la mobilité occupationnelle des individus au cours de 5 ans de carrière. En effet, différents travaux suggèrent que l'origine sociale n'a pas un effet forcément décroissant au cours du cycle de vie et qu'elle peut continuer à influencer le déroulement de la carrière professionnelle, même après qu'elle ait exercé une influence sur l'éducation et le premier emploi. L'article décrit tout d'abord la mobilité occupationnelle au Pérou en tant qu'instrument utile pour comprendre la dynamique du marché du travail péruvien et met en avant l'importance de cette mobilité et sa complexité. Ensuite, il cherche à tester si l'origine sociale est déterminante dans l'évolution de la carrière, même après avoir contrôlé ses effets sur les étapes antérieures du cycle de vie.

**Mots clés :** Mobilité occupationnelle, égalité d'opportunités, Pérou, probit bivarié

**JEL :** J62, D63.

## ABSTRACT

The Peruvian society is characterized by a highly unequal income distribution. But, is it also an unfair society in accordance with the criterion of equality of opportunities? Several studies have shown that social origin does not necessarily have a decreasing effect during the life cycle. For example, social origin can continue to influence the professional career even after it has exerted an influence on education and the first job. This paper tries to answer this question by analysing the effect of social origin on occupational mobility in Peru. It describes occupational mobility in Peru as a useful instrument to understand the dynamics of the Peruvian labor market and shows the importance and complexity of this mobility. Finally, it attempts to explain to what extent social background affects occupational mobility after controlling for its effect on the prior stages of the life cycle.

**Key words :** Occupational mobility, equality of opportunities, Peru, Bivariate probit model

**JEL Code :** J62, D63.

---

<sup>1</sup> Les auteurs tiennent à remercier Denis Cogneau, Philippe De Vreyer, Javier Herrera et François Roubaud pour leurs précieux commentaires.

## Table des matières

<b>INTRODUCTION.....</b>	<b>5</b>
<b>1. ETUDES PRECEDENTES, PRESENTATION DES DONNEES, DU CONTEXTE ECONOMIQUE ET CONSTRUCTION DES GROUPES OCCUPATIONNELS.....</b>	<b>8</b>
1.1. Etudes précédentes sur la mobilité occupationnelle intragénérationnelle au Pérou.....	8
1.2. Les données.....	8
1.3. Le contexte économique entre 1985 et 1990.....	9
1.4. La construction des groupes occupationnels : l'analyse par cluster.....	10
<b>2. LA MOBILITE OCCUPATIONNELLE A LIMA.....</b>	<b>12</b>
2.1. L'ampleur de la mobilité.....	12
2.2. Les matrices de transition.....	12
2.3. Les flux entre les groupes occupationnels entre 1982 et 1990.....	13
<b>3. LES EFFETS DE L'ORIGINE SOCIALE SUR LA MOBILITE OCCUPATIONNELLE.....</b>	<b>15</b>
<b>3.1. Le lien entre l'origine sociale et la mobilité occupationnelle par l'analyse descriptive.....</b>	<b>15</b>
3.1.1 L'origine sociale influence-t-elle la mobilité occupationnelle en terme de mouvements ? .....	15
3.1.2 L'origine sociale influence-t-elle la mobilité occupationnelle en terme de trajectoires ?.	16
<b>3.2. Le lien entre l'origine sociale et la mobilité occupationnelle par l'analyse économétrique.....</b>	<b>17</b>
3.2.1 Test de l'hypothèse de persistance de l'effet de l'origine sociale au cours de la carrière par les modèles log-linéaires.....	17
3.2.1.1 Les modèles.....	18
3.2.1.2 Les résultats.....	19
3.2.2 Test de l'effet de l'origine sociale sur la carrière professionnelle par le modèle probit bivarié.....	20
3.2.2.1 Le choix du modèle approprié.....	20
3.2.2.2 Le choix d'un instrument.....	22
3.2.2.3 L'estimation du modèle.....	23
<b>CONCLUSION.....</b>	<b>25</b>
<b>REFERENCES BIBLIOGRAPHIQUES.....</b>	<b>27</b>
<b>ANNEXES.....</b>	<b>29</b>
Annexe 1 : Les limites du panel.....	29
Annexe 2 : La construction de la classification des catégories occupationnelles.....	31
Annexe 3 : Présentation des modèles log-linéaires.....	34
Annexe 4 : Tests de spécification des modèles log-linéaires.....	37
Annexe 5 : Estimations des modèles 1 et 2.....	39

## Liste des tableaux

Tableau 1 :	Description de la mobilité à Lima M. entre 1985 et 1990 (en %)	12
Tableau 2 :	Les transitions entre les catégories d'occupation à Lima M., 1985-1990	13
Tableau 3 :	Nombre de changements d'occupation entre 1982 et 1990 selon l'origine sociale (%) Lima M.	16
Tableau 4 :	Matrices de transition entre 1982 et 1990 en fonction de l'origine sociale	17
Tableau 5 :	Modèles log-linéaires avec l'origine sociale, l'occupation en 1985, l'occupation en 1990	19
Tableau 6 :	Résultats des estimations à partir des modèles probit bivariés	24

## Liste des graphiques

Graphique 1 :	Evolution du taux de croissance du PIB entre 1980 et 1991 (%)	10
Graphique 2 :	Transitions entre les groupes occupationnels entre 1982, 1985 et 1990 (%)	14

## INTRODUCTION

Une société ne connaît la cohésion sociale que si elle est considérée par le plus grand nombre de ses membres comme étant une société juste. Généralement, plus une société a une répartition égalitaire des revenus, plus elle est considérée comme juste. Or, comme le rappelle Stiglitz (2000), le critère d'égalité de revenus n'est pas le seul critère permettant de définir si une société est juste ou non. Pour deux sociétés ayant le même degré d'égalité de revenus, la première pourra être regardée comme étant plus juste que la seconde si l'égalité des chances y est relativement mieux assurée c'est-à-dire si chacun de ses membres dispose des mêmes chances d'obtenir une certaine position dans l'échelle sociale, alors que dans la seconde société, la position atteinte est héritée. La société péruvienne, comme beaucoup d'autres sociétés d'Amérique latine, est caractérisée par une répartition très inégalitaire des revenus. Cela en fait-elle pour autant une société injuste au regard du critère d'égalité des chances ? Si c'est le cas, c'est-à-dire si l'égalité des chances y est faible, alors l'origine sociale d'un individu va influencer son statut social à différents stades de son cycle de vie.

Comme l'ont montré de nombreux auteurs, l'influence de l'origine sociale sera dans un premier temps déterminante dans l'acquisition d'éducation<sup>2</sup>, notamment *via* l'imperfection des marchés de capitaux, la transmission intergénérationnelle des aptitudes, ou encore en modulant la motivation à étudier. Jacoby (1994) a testé empiriquement pour le Pérou l'argument développé par Becker et Tomes (1986)<sup>3</sup> selon lequel l'imperfection des marchés des capitaux conduit à un sous-investissement des familles contraintes financièrement dans l'éducation de leurs enfants. Il montre qu'en effet, au Pérou, les enfants de familles contraintes financièrement ont un parcours scolaire plus lent que ceux de familles non contraintes.

L'origine sociale est déterminante ensuite dans l'accès au premier emploi, pour un niveau d'études donné (Müller et Shavit, 1998 ; Coury, 2000), principalement de par le tissu de relations sociales et les informations sur le fonctionnement et les opportunités du marché du travail qu'elle procure (Goux et Maurin, 1998). Rodriguez (1995) montre ainsi, à l'aide d'entretiens qualitatifs, que le rôle de la famille et des amis est déterminant à Lima pour obtenir un emploi qualifié après une formation supérieure, les jeunes diplômés sans soutien familial ou relationnel ayant un risque très élevé de ne pas exercer la profession pour laquelle ils ont été formés. L'origine sociale enfin joue un rôle dans le choix d'une carrière professionnelle (Hout et Rosen, 1999).

Ainsi, être né dans une famille au statut social élevé confère de nombreux avantages dans l'accès à l'éducation ou dans l'accès au premier emploi. Même s'il existe un consensus selon lequel l'action de la position sociale n'est pas unique au cours du parcours scolaire mais répétitive et cumulative (Gravot, 1995), une idée généralement acceptée est qu'une fois que le premier emploi est déterminé, l'influence de la famille diminue avec le temps, s'estompant au profit des qualités propres de la personne, de son expérience et du hasard. Pourtant, d'autres hypothèses font penser que cette influence perdure.

La persistance du rôle de l'origine sociale sur l'évolution de la carrière professionnelle peut s'expliquer tout d'abord par le fait que l'origine sociale conditionne la motivation à connaître une mobilité ascendante, comme le souligne Piketty (1998). Selon cet auteur, un individu a un statut élevé si les autres considèrent qu'il est doué. Cependant, ils n'ont aucun moyen d'estimer les talents de l'individu autre que de constater s'il réussit professionnellement ou non. L'individu est donc incité à

---

<sup>2</sup> Pour une revue de littérature sur le sujet, se référer à Haveman & Wolfe (1995).

<sup>3</sup> Dans le modèle de Becker & Tomes (1986), sous l'hypothèse de marchés de capitaux imparfaits, les parents ne peuvent pas s'endetter. Pour pouvoir investir en capital humain ils sont donc obligés soit à vendre des biens, soit à réduire leur consommation. Les parents les plus pauvres, qui ne disposent pas de biens à vendre, auront donc tendance à diminuer leurs investissements en capital humain car réduire leur niveau de consommation revient à diminuer leur niveau d'utilité. Par conséquent, les investissements en capital humain des familles pauvres ne dépendent pas seulement des capacités de l'enfant et des dépenses publiques, mais aussi de la richesse des parents et de leur générosité. Les familles pauvres investiront en capital humain tant que la baisse de l'utilité marginale, engendrée par la réduction du niveau de consommation nécessaire au financement de dépenses supplémentaires en capital humain, est plus faible que la hausse de l'utilité marginale générée par l'accroissement de l'utilité de l'enfant. Cette hausse est liée à celle du revenu du travail à la suite de l'investissement supplémentaire en capital humain. Tout ceci dépendra du niveau du revenu familial.

fournir des efforts afin d'avoir une bonne profession -ou un bon revenu- pour prouver aux autres qu'il a du talent. Dans son modèle, Piketty s'inspire de la théorie sociologique du groupe de référence (Merton 1953, Boudon 1973) et montre que dans les sociétés où les qualités d'une personne sont estimées par les autres selon qu'elle a ou non réussi professionnellement, l'origine sociale a un rôle persistant sur la carrière des individus en modelant la motivation à acquérir un statut déterminé. En effet, les classes sociales élevées ont une très grande motivation à atteindre une position élevée car elles font face à de fortes pressions sociales afin de ne pas déroger à leur rang. Les classes sociales les plus défavorisées pensent en revanche que leurs chances d'ascension sont faibles, du fait d'échecs passés, et considèrent que l'effort à fournir pour monter dans l'échelle sociale est trop coûteux par rapport au gain escompté. Cet argument rejoint celui d'Akerlof (1997) selon lequel prendre une décision qui éloigne de son entourage coûte cher en terme d'utilité car elle peut nuire aux relations avec ses proches. Ainsi, une forte mobilité ascendante a un impact important sur le réseau social de chacun et cet impact est le premier déterminant de la décision de fournir ou non un effort pour s'élever dans l'échelle sociale. Les classes plus défavorisées sont donc selon cette hypothèse moins motivées dans la recherche d'une mobilité sociale fortement ascendante.

Les propos tenus par Carlos, ouvrier qualifié en mécanique, âgé de 30 ans, vivant dans un quartier relativement défavorisé de Lima (Comas) viennent corroborer ces hypothèses<sup>4</sup> :

(58) « *Pour le moment, je ne veux pas me sentir [d'un niveau] supérieur à [celui de] mon père. Je ne sais pas si je réussirai ou non mais je ne veux pas me sentir [d'un niveau] supérieur parce que j'ai peur qu'il se sente mal à l'aise, que je le fasse se sentir mal à l'aise.* »<sup>5</sup>

La persistance du rôle de l'origine sociale sur l'évolution de la carrière professionnelle peut s'expliquer ensuite par la transmission intergénérationnelle du capital social (Granovetter, 1974). Plusieurs études ont souligné le rôle du capital social lors de l'accès à l'emploi dans les pays en développement (Kouamé et Guyé, 2000 ; Cohen et House, 1996 ; Bocquier, 1996). Bien qu'il y ait eu peu de tests empiriques sur l'importance du capital social dans l'accès aux emplois dans les pays en développement, on admet généralement que dans ces pays, l'information est divulguée par des canaux informels, dans la mesure où les journaux occupent une place mineure dans les modes de recrutement et que les services de placements officiels sont quasi-inexistants (BID, 2004 pour l'Amérique Latine). L'accès de chacun aux informations sur les opportunités d'emplois et aux employeurs dépend donc de son réseau social (Coury, 2000). Or la possession de tels réseaux est fortement corrélée à l'origine sociale dans la mesure où les parents peuvent soit transmettre à leurs enfants une partie de leur capital social, soit les aider à créer le leur. Par exemple, au Pérou, venir d'une famille au niveau de vie élevé permet d'entrer dans des écoles privées sélectives où se côtoient des jeunes du même milieu social. Les écoles et surtout les universités sont un centre important de socialisation et c'est souvent là que se constitue une grande partie du capital social. Une origine sociale élevée confère donc un avantage supplémentaire dans l'évolution de la carrière, celui d'être doté d'un capital social important.

Enfin, un troisième canal, certes plus indirect, par lequel l'origine sociale peut continuer à influencer l'évolution de la carrière professionnelle est celui du capital humain. En effet, certains auteurs ont montré que les avantages des diplômés ne se manifestent pas seulement au moment de leur entrée sur le marché du travail mais également plus tard dans leur vie active, en facilitant leurs promotions et par conséquent une mobilité professionnelle ascendante (Galor et Sicherman, 1990 ; Sicherman, 1990). Bocquier (1996) confirme ces résultats pour le Sénégal en montrant qu'à Dakar, le capital humain continue d'influer sur la carrière professionnelle tout au long de la vie active. Or, comme il a été mentionné précédemment, l'origine sociale a un rôle non négligeable dans la détermination du niveau de capital humain.

Ces différents arguments suggèrent donc que l'origine sociale n'a pas un effet forcément décroissant au cours du cycle de vie et qu'elle peut continuer à agir sur le déroulement de la carrière

---

<sup>4</sup> Propos recueillis lors d'un travail de collecte d'une centaine de récits de vie, réalisé à Lima en 2003 par L. Pasquier-Doumer et visant à connaître les canaux de mobilité sociale à Lima.

<sup>5</sup> "Por el momento no me quiero sentir más. No sé si tendré o no tendré pero no quiere sentirme más porque tengo miedo de que mi padre se sienta incómodo, yo hacerlo sentir incómodo."

professionnelle, même après qu'elle ait exercé une influence sur le premier emploi. Elle peut agir durablement sur la carrière professionnelle directement en conditionnant la motivation à acquérir un statut social et en déterminant le capital social et indirectement *via* le capital humain. C'est sous cet angle bien précis que cet article se propose d'étudier l'égalité d'opportunités au Pérou, à savoir si l'effet de l'origine sociale persiste au cours de la carrière professionnelle. Toutefois, si c'est le cas, on ne cherchera pas à déterminer par quels canaux cet effet perdure.

Cette problématique est apparue relativement tôt dans la littérature sociologique puisqu'en 1967 Blau et Duncan affirmaient :

*« Le statut occupationnel du père n'influence pas seulement la carrière du fils via son rôle dans l'éducation et le premier emploi de ce dernier. Il a également un effet retardé qui persiste une fois que sont contrôlées statistiquement les différences de niveaux d'éducation et d'expérience lors des premières années de carrière »*<sup>6</sup> (p.403).

Cependant très peu de vérifications empiriques ont été menées et aucune pour un pays en développement. Goux et Maurin (1997) testent empiriquement cette hypothèse pour la France à l'aide de modèles log-linéaires. Ils concluent que l'origine sociale, mesurée par l'activité du père, est déterminante non seulement en début de carrière mais aussi tout au long de la carrière : parmi les personnes de même qualification et avec le même début de carrière, les personnes d'origine sociale favorisée ont souvent une carrière plus réussie que les personnes d'origine sociale modeste. Ils montrent également que l'effet du niveau d'éducation atteint, même s'il influe sur l'ensemble du parcours professionnel, est déterminant surtout en début de carrière car son importance relative décroît au cours de la carrière au profit de l'effet de l'origine sociale.

Cependant, le résultat peut être sensible au choix de la méthodologie et à notre connaissance, il n'existe aucune vérification empirique se basant sur une méthodologie différente.

L'enjeu de cet article est double. Il cherche d'une part à décrire la mobilité occupationnelle au Pérou afin d'améliorer la connaissance générale sur le fonctionnement du marché du travail péruvien et, en particulier, de mieux appréhender sa dynamique. Il tente d'autre part de répondre à la question suivante : quel est le rôle de l'origine sociale sur la mobilité occupationnelle ?

Pour répondre à ce double enjeu, nous disposons de données assez exceptionnelles pour un pays en développement, c'est-à-dire des données de panel pour la période 1985 à 1990 qui contiennent des informations sur l'origine sociale de l'enquêté et des historiques d'emploi.

Dans une première partie, nous présentons les données, le contexte économique de la période ainsi que la méthode par laquelle nous avons construit les groupes occupationnels. Dans une deuxième partie, nous décrivons la mobilité occupationnelle à Lima. Nous montrons que la mobilité occupationnelle a été de grande ampleur sur la période et les transitions entre les groupes d'occupation très complexes. Enfin, dans une troisième partie, nous analysons l'influence de l'origine sociale sur la mobilité occupationnelle. Cette analyse, qui confirme que l'origine sociale a un effet sur ce type de mobilité, ne parvient pas à montrer formellement la persistance de cet effet au cours du temps, du fait des importantes difficultés techniques que soulève cette problématique mais présente un certain nombre de constatations qui laissent à penser qu'une telle persistance existe.

---

<sup>6</sup> "Father's occupational status, on the other hand, not only influences son's career achievements by affecting his education and first job, but it also has a delayed effect on achievements that persists when differences in schooling and early career experience are statistically controlled".

# 1. ETUDES PRECEDENTES, PRESENTATION DES DONNEES, DU CONTEXTE ECONOMIQUE ET CONSTRUCTION DES GROUPES OCCUPATIONNELS

## 1.1. Etudes précédentes sur la mobilité occupationnelle intragénérationnelle au Pérou

Au Pérou, deux études se sont intéressées à la mobilité occupationnelle<sup>7</sup> à Lima Métropolitaine. La première met en évidence l'importance de ce type de mobilité pour la période 1963-1973 ainsi que son caractère sélectif (Suarez, 1975). L'auteur constate qu'il existe des différences selon le sexe -les hommes sont plus mobiles que les femmes-, les groupes d'âge -plus de mobilité chez les jeunes adultes-, les parcours migratoires -les plus mobiles sont ceux qui ont récemment migré à Lima- et le niveau d'études -les individus qui ont fait des études supérieures ont des probabilités plus fortes de connaître une mobilité ascendante-. La seconde étude analyse le problème de la mobilité à partir de deux enquêtes emploi réalisées respectivement en 1989 et 1991 (Chacaltana, 1993). Chacaltana observe une mobilité occupationnelle relativement forte à Lima qui s'accroît entre 1989 et 1991. Cet accroissement s'expliquerait par les premiers effets de la réforme structurelle, appliquée par le gouvernement Fujimori en 1990.

La caractéristique essentielle de ces deux études est qu'elles ont été réalisées à partir des historiques de l'emploi des enquêtes utilisées. Par conséquent, ce ne sont pas des données dynamiques de type panel mais des informations *ex-post* qui dépendent de la bonne ou de la mauvaise mémoire des enquêtés<sup>8</sup>. Surtout, aucune de ces deux études n'introduit l'origine sociale comme facteur explicatif de ce type de mobilité.

## 1.2. Les données

Si le sujet de la mobilité occupationnelle intragénérationnelle a été peu traité, c'est qu'il est très exigeant en terme de données requises. La mobilité occupationnelle ne peut être étudiée qu'à partir de données dynamiques, soit de type panel soit à partir d'historiques de l'emploi. De plus, pour tester l'effet de l'origine sociale sur la mobilité occupationnelle, il faut disposer d'informations sur les parents de l'enquêté, informations qui sont rares, en particulier dans les pays en développement.

Ces deux conditions sont remplies au Pérou puisque nous disposons d'un panel pour la période 1985-1990 construit à partir d'un échantillon de Liméniens enquêtés en 1985/86 et en 1990 et pour lesquels sont connues l'activité et l'éducation des parents. Les personnes ont de plus été enquêtées sur leur activité professionnelle précédant celle qu'elles exerçaient en 1985 et en 1990.

Le panel a été construit à partir de deux enquêtes sur les niveaux de vie, la ENNIV 1985/86 et la ENNIV 1990, élaborées suivant la méthodologie LSMS<sup>9</sup>. Ce type d'enquêtes a été développé par la Banque mondiale avec l'objectif de générer de l'information sur le bien-être et sur la qualité de vie des ménages des pays en développement. Les premières LSMS furent mises en place en 1985 en Côte d'Ivoire et au Pérou<sup>10</sup>. Celle du Pérou fut réalisée conjointement avec l'Institut CUANTO et, depuis,

---

<sup>7</sup> Il s'agit ici de la mobilité occupationnelle intragénérationnelle, c'est-à-dire la mobilité que connaît une personne entre deux périodes de sa carrière professionnelle. Elle est à distinguer de la mobilité intergénérationnelle qui définit la mobilité entre deux générations, c'est-à-dire entre le statut professionnel des parents et celui de leur enfant. Pour une étude sur la mobilité occupationnelle intergénérationnelle au Pérou, se référer à Benavides (2002).

<sup>8</sup> Grimm et Bonneuil (2001) présentent une revue de littérature sur ce problème de mémoire. Les auteurs se basent sur un travail de Auriat (1996) et indiquent à propos des événements renouvelables que : « les expériences en psychologie cognitive ont montré que ce type d'évènement crée un souvenir générique ayant pour caractéristique une meilleure mémorisation des événements se situant au début et à la fin de la série que de ceux se situant au milieu. De plus la vie professionnelle est considérée comme un thème plus sensible que la mobilité résidentielle ou la constitution familiale. Les gens ont tendance à ne pas déclarer de longues périodes d'inactivité, qu'ils considèrent comme des périodes antisociales ». Grimm et Bonneuil citent aussi le travail de Thélot (1990) et indiquent que cet auteur : « a essayé de séparer l'erreur volontaire de l'erreur de mémoire et de quantifier cette dernière, concernant les réponses sur la mobilité professionnelle sondée par l'Enquête Emploi de l'INSEE. L'auteur constate que l'erreur de mémoire constitue entre 0.5 % et 35 % de l'ensemble des erreurs ». En outre, selon Grimm et Bonneuil, d'autres études montrent que l'ampleur de l'erreur temporelle (1 semaine, 1 mois 1 an) n'est pas directement liée à la longueur de la période rétrospective mais au type d'évènement concerné et que les réponses dans les enquêtes rétrospectives sont relativement peu fiables pour les événements qui concernent le changement d'activité.

<sup>9</sup> « *Living Standard Measurement Survey* »

<sup>10</sup> Ces enquêtes ainsi que les informations qui les concernent sont disponibles sur le site Internet de la Banque mondiale : <http://www.worldbank.org/lsm/>.

elle est disponible pour les années 1985/86, 1990, 1991, 1994, 1996, 1997 et 2000<sup>11</sup>. L'enquête ENNIV 1985/86 se déroula entre juillet 1985 et juillet 1986 et son échantillon est national<sup>12</sup>. Cette enquête est constituée de 27.445 individus appartenant à 4.913 ménages et comprend un volet sur l'éducation et l'activité des parents de l'enquêté. L'enquête ENNIV 1990 - menée entre juin et juillet 1990- ne concerne que Lima Métropolitaine. Cette enquête est constituée de 7.720 individus appartenant à 1.525 ménages. Elle comporte un module de type panel qui permet de retrouver 727 ménages enquêtés en 1985/86, soit 3.335 individus. Après apurement des données, différents tests -par sexe, âge et date de naissance- ont été effectués pour s'assurer de la qualité du panel. Ces tests ont conduit à retenir dans le panel 709 ménages. Seuls les individus de plus de 25 ans<sup>13</sup> et de moins de 65 ans<sup>14</sup> sont conservés dans cette analyse, soit 1.207 individus.

Les individus du panel ne représentent que 43 % de l'échantillon total des individus âgés de 26 à 64 ans de la ENNIV de 1985/86 vivant à Lima (cf. tableau A, annexe 1). Cette différence ne s'explique pas par l'attrition du panel puisque l'ENNIV 1990 marque la constitution de celui-ci. On peut cependant redouter un biais de sélection, i.e. que les individus présents dans l'échantillon panel aient des caractéristiques différentes des autres individus de l'enquête de 1985. Par conséquent, avant d'utiliser le panel nous avons évalué les distorsions potentielles liées à une mauvaise sélection de l'échantillon du panel. Pour ce faire, nous avons testé la significativité des différences entre les caractéristiques des individus du panel et celles du reste des individus de l'échantillon total des Liméniens en 1985. On observe que les individus du panel ont des caractéristiques relativement différentes (cf. annexe 1). En particulier, ils semblent avoir une origine sociale plus défavorisée, notamment quant à la dotation en capital humain de leurs parents.

Grâce à l'historique de l'emploi de l'enquête de 1985, il a été possible de connaître l'activité professionnelle qu'exerçaient les individus de notre échantillon en 1982. L'année 1982 a été retenue car elle vérifie les deux critères suivants. D'une part, elle est suffisamment éloignée de 1985 pour qu'il y ait une certaine variabilité entre l'occupation de 1982 et celle de 1985. D'autre part, elle n'est pas trop éloignée dans la mesure où la part des individus qui ont changé plus d'une fois d'activité entre 1982 et 1985 -et donc pour lesquels on perd l'information sur leur occupation en 1982- est relativement faible (7 % de l'échantillon). Enfin, le fait qu'elle n'est pas trop éloignée de 1985 réduit la probabilité d'erreur de mémoire

### **1.3. Le contexte économique entre 1985 et 1990**

La période d'analyse se situe pendant le gouvernement populiste d'Alan García (juillet 1985-avril 1990). Elle peut être divisée en deux sous-périodes (cf. graphique 1). La première, 1985-1987, est caractérisée par une politique de fortes dépenses publiques menant notamment à une multiplication des emplois publics<sup>15</sup> et à une hausse des salaires. Ces dépenses publiques, accompagnées d'un gel des prix et d'une diminution du paiement de la dette extérieure<sup>16</sup> visaient à réactiver l'économie en stimulant la demande. Cette politique a d'abord porté ses fruits puisqu'elle s'accompagne jusqu'en 1987 d'une croissance soutenue<sup>17</sup> mais également d'un important déficit public.

---

<sup>11</sup> La ENNIV de 2000 fut réalisée uniquement par CUANTO.

<sup>12</sup> Il manque cependant les départements de Ayacucho, Apurimac et Huancavelica. A l'époque ces trois départements, qui représentaient 6 % de la population totale du pays, étaient considérés comme dangereux à cause du terrorisme.

<sup>13</sup> Cette limite d'âge inférieure a été fixée afin de s'assurer que ceux qui ont suivi des études supérieures les ont achevées et qu'ils sont entrés sur le marché du travail. On évite ainsi de prendre en compte les jeunes qui poursuivent leurs études en 1985 et qui occupent des « petits boulots » pour les financer. On surestimerait alors la mobilité de ces jeunes entre 1985 et 1990 car ces « petits boulots » sont en général très différents de l'occupation pour laquelle ils se forment et qu'ils occupent à la fin de leurs études.

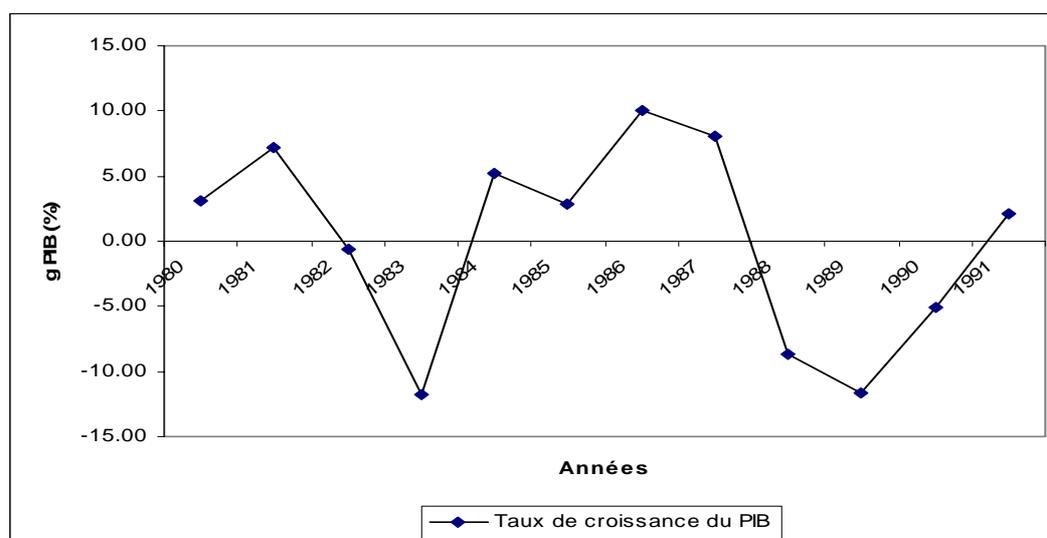
<sup>14</sup> En effet, inclure ceux qui travaillent encore après 65 ans générerait un biais de sélection dans la mesure où ces personnes ont un profil bien particulier.

<sup>15</sup> Entre 1985 et 1987, on assiste à la création de plus de 130.000 emplois publics (Herrera, 1999)

<sup>16</sup> Le gouvernement décida de limiter le remboursement de la dette extérieure aux 10% des revenus procurés par les exportations. Cette décision conduisit le FMI à exclure le Pérou du système financier international.

<sup>17</sup> Le taux de croissance du PIB par tête est en moyenne de plus de 6 %.

**Graphique 1 : Evolution du taux de croissance du PIB entre 1980 et 1991 (%)**



Source : INEI, construction des auteurs.

La seconde sous-période, 1988-1990, est celle d'une grave crise économique au cours de laquelle le Pérou connaît un taux de croissance négatif, autour de 10 % en 1988, et qui conduit à une politique de restriction budgétaire. La crise économique s'accompagne d'un processus d'hyperinflation<sup>18</sup> comme n'en a jamais connue le pays (Herrera 1999) et d'une chute des salaires réels de 75% en moyenne.

En décomposant la population en déciles selon le niveau de consommation, Glewwe et Hall (1994) montrent qu'à Lima, sur la période 1985-1990, chaque décile connaît une baisse de leur consommation réelle totale de plus de 50 % et que les deux déciles les plus pauvres sont les plus touchés. De plus, l'évolution de l'indice de Gini pendant cette période montre une hausse importante des inégalités<sup>19</sup> (Herrera, 1999).

#### **1.4. La construction des groupes occupationnels : l'analyse par cluster**

L'approche qui a été retenue ici est l'analyse de la mobilité entre groupes occupationnels. Cette approche a été préférée à celle de la mobilité le long d'une variable continue, comme le revenu ou la consommation du ménage pour plusieurs raisons. La première raison est que l'occupation est un meilleur indicateur du statut social. En effet, comme le rappelle Weber, le statut social est multidimensionnel<sup>20</sup>. L'occupation est une approximation plus riche du statut social qu'une variable continue comme le revenu car elle inclut, outre la notion de niveau de vie, celle de prestige voire de pouvoir et qui plus est, donne une indication sur le niveau d'éducation atteint. La seconde raison est que l'occupation est un meilleur indicateur du revenu permanent que le revenu ou la consommation à un point donné du temps (Zimmermann, 1992). Enfin, la dernière raison, d'ordre empirique, est que nous ne disposons pas de bonne variable continue. Les deux variables candidates sont les revenus du travail de la personne et le niveau de consommation du ménage auquel elle appartient entre 1985 et 1990. Elles sont cependant très peu fiables dans l'enquête de 1990 du fait de l'hyper-inflation qui a marqué cette année<sup>21</sup>. Un traitement statistique relativement lourd aurait certes pu améliorer la qualité

<sup>18</sup> Le taux annuel d'inflation fut de 3.400 % en 1989.

<sup>19</sup> Il passe à Lima Métropolitaine de 0,37 en 1985 à 0,43 en 1990.

<sup>20</sup> Selon Weber, le statut social comprend trois dimensions : la richesse économique, le pouvoir et le prestige.

<sup>21</sup> En effet, l'enquête a été réalisée entre juillet et septembre de l'année 1990 et en août a eu lieu un des chocs économiques les plus importants qu'a connu le Pérou qui a eu pour conséquence une inflation de plus de 200 % entre début juillet et fin août. La répercussion de ce choc sur les revenus du travail a été considérable et n'a pas été la même pour tous les individus, pour ceux enquêtés en juillet que pour ceux enquêtés fin août, pour les salariés que pour les entrepreneurs. La mobilité de revenus que l'on peut constater entre 1985 et 1990 est donc biaisée si l'on ne procède pas à un lourd travail de redressement des données. L'hyper-inflation a eu un effet encore plus fort sur la qualité de l'estimation de la consommation. Les prix ont été tellement instables durant la période de l'enquête que les ménages ont pu oublier le prix des produits qu'ils ont consommés deux mois auparavant, ce prix ayant depuis changé plusieurs dizaines de fois. Ces biais –différence entre ceux interrogés en juillet et ceux interrogés en septembre et mauvaise mémoire du fait de la forte variabilité des prix- ne peuvent être corrigés (Glewwe et Hall, 1994).

de ces variables mais l'enjeu théorique ne nous a pas semblé en valoir la peine même si utiliser une variable continue aurait présenté deux avantages. Le premier est celui d'éviter un effet de frontière, commun à toute étude se basant sur des groupes d'occupations. Un individu qui se trouve proche de la frontière droite<sup>22</sup> d'un des groupes d'occupations et qui connaît une très faible mobilité ascendante sera considéré comme mobile alors qu'un individu proche de la frontière gauche<sup>23</sup> qui vit une très forte mobilité ascendante pourra être considéré comme immobile. Or, la position au sein du groupe d'occupations en 1985 dépend de l'origine sociale. Le second avantage est d'éviter l'hypothèse de la permanence des groupes occupationnels entre 1985 et 1990, c'est-à-dire qu'appartenir à un certain groupe procure le même statut social en 1985 qu'en 1990, relativement aux autres groupes. Cette hypothèse est discutée dans l'annexe 2.

Maintenant que nous avons explicité notre choix d'étudier la mobilité entre groupes occupationnels plutôt que le long d'une variable continue, il nous faut préciser la façon dont nous avons construit ces groupes. Les choix méthodologiques pour la construction des groupes sont importants puisque l'incidence ainsi que le type de mobilité (ascendante ou descendante) en dépendent fortement. La classification retenue doit répondre à deux critères. Premièrement, elle doit être suffisamment fine pour pouvoir observer le plus de transitions individuelles possibles entre les groupes occupationnels. Une classification trop agrégée sous-estimerait la mobilité. Deuxièmement, elle doit permettre une interprétation de la mobilité en termes hiérarchiques, c'est-à-dire en termes de mobilité ascendante ou descendante.

L'occupation des enquêtés est classifiée, dans les deux enquêtes utilisées, suivant la classification internationale des professions élaborée par l'OIT<sup>24</sup>. Afin d'obtenir une classification plus agrégée et hiérarchisée, nous avons utilisé l'approche de la classification ascendante hiérarchique (« *cluster analysis* »). Cette approche (présentée dans l'annexe 2) permet d'associer des professions dans un groupe de sorte que les professions soient les plus similaires au sein de ce groupe et les plus différentes de celles des autres groupes. Les critères de hiérarchisation retenus sont le niveau de vie que procure chaque occupation et le niveau de qualification qu'elle requiert<sup>25</sup>. Le niveau de vie est approché par trois variables : le niveau moyen et le niveau médian de la consommation mensuelle du ménage par tête et le pourcentage de ménages épargnant au sein de chaque occupation. Le niveau de qualification associé à chaque occupation est le nombre moyen d'années d'éducation des personnes exerçant cette occupation. Nous avons ainsi constitué trois groupes (présentés en détail dans l'annexe 2), à partir de l'enquête de 1985<sup>26</sup> :

- le groupe 1 regroupant les occupations procurant le niveau de vie le plus élevé et au niveau de qualification le plus élevé,
- le groupe 2, groupe intermédiaire selon ces deux critères et
- le groupe 3, regroupant les occupations les plus précaires (les niveaux de vie et de qualification moyen au sein du groupe 3 sont deux fois plus faibles que ceux du groupe 1).
- Le groupe 1 est plus éloigné du groupe 2 que le groupe 2 ne l'est du groupe 3.

---

<sup>22</sup> C'est-à-dire que cet individu a une occupation très proche du groupe occupationnel supérieur à celui où il a été classé, les groupes occupationnels étant hiérarchisés.

<sup>23</sup> C'est-à-dire que cet individu a une occupation très proche du groupe occupationnel inférieur à celui où il a été classé.

<sup>24</sup> « Clasificación Internacional Uniforme de Ocupaciones », (CIDO, rev.70)

<sup>25</sup> On a préféré cette approche à une approche en terme de classe comme celle définie par Goldthorpe (1980) pour la raison suivante. Le schéma de classe défini par cet auteur est basé sur les relations au travail, en particulier sur la distinction entre ceux qui achètent le travail des autres, ceux qui n'achètent pas le travail des autres mais qui ne vendent pas le leur et ceux qui vendent leur travail. Ce schéma a un sens lorsque l'on s'intéresse à la mobilité intergénérationnelle mais en a beaucoup moins dans le cadre de la mobilité intragénérationnelle où le passage d'une classe à l'autre a une très faible occurrence.

<sup>26</sup> En ne retenant, par souci de cohérence, que les personnes vivant à Lima âgées de 26 à 64 ans.

## 2. LA MOBILITE OCCUPATIONNELLE A LIMA

Avant d'étudier l'influence de l'origine sociale sur la mobilité occupationnelle entre 1985 et 1990, il est indispensable d'étudier au préalable quelle a été la mobilité occupationnelle pendant la période. Autrement dit, nous voulons observer l'ampleur et le sens des transitions individuelles entre les groupes d'occupation préalablement définis.

### 2.1. L'ampleur de la mobilité

Le tableau 1 permet d'observer qu'en cinq ans, 29 % des Liméniens ont changé de situation professionnelle transitant d'un groupe d'occupation vers un autre<sup>27</sup>. La proportion de mobiles ascendants a été significativement plus importante que celle des mobiles descendants (19 % contre 10 %). Dans les deux cas, les transitions vers le groupe le plus proche ont été les plus importantes. Enfin, les hommes se distinguent des femmes uniquement par un taux d'immobilité légèrement plus élevé<sup>28</sup>.

**Tableau 1 : Description de la mobilité à Lima M. entre 1985 et 1990 (en %)**

	1985-1990		
	Ensemble	Hommes	Femmes
<b>Inmobiles</b>	71,1	71,8	69,7 **
<b>Mobiles</b>			
- <b>Part des mobiles ascendants</b>	18,6	18,2	19,1
Saut d'un groupe	16,4	15,3	18,5
Saut de deux groupes	2,1	2,9	1,0
- <b>Part des mobiles descendants</b>	10,5	10,0	11,2
Saut d'un groupe	9,5	9,0	10,3 *
Saut de deux groupes	0,9	1,0	0,9
<b>Total</b>	100	100	100
<b>N. observations</b>	810	495	315

Source : Panel LSMS 1985-1990, construction des auteurs.

Champ : Les individus de 26-64 ans en 1985 exerçant une activité en 1985 et 1990.

Note : Tests de différences entre hommes et femmes. \* différence significative au seuil de 10 %, \*\* de 5 %, \*\*\* de 1 %

### 2.2. Les matrices de transition

Les matrices de transitions (cf. tableau 2) sont un outil pertinent pour décrire le sens des différentes transitions entre les groupes d'occupation. Elles montrent que les individus les moins mobiles exercent des professions intermédiaires (groupe 2) alors que les plus mobiles sont ceux du groupe le plus défavorisé (groupe 3). De plus, les individus des groupes 2 et 3 ont peu de chances de transiter vers le groupe le plus favorisé (groupe 1) et inversement, les personnes des groupes 1 et 2 ont une faible probabilité de transiter vers le groupe le plus défavorisé.

On peut observer des différences entre les types de transitions des hommes et celui des femmes. En général, les femmes semblent avoir moins d'opportunités que les hommes de connaître une forte mobilité car elles transitent majoritairement entre les groupes 2 et 3. En effet, les femmes appartenant au groupe 2 en 1985 ont moins de chances que les hommes de transiter vers le groupe 1 et plus de chances de transiter vers le groupe 3. De même, les femmes du groupe 3 transitent plus souvent que leurs équivalents hommes vers le groupe 2. Enfin, la part des immobiles du groupe 1 et du groupe 3 est plus élevée chez les femmes.

<sup>27</sup> Cette proportion est de 55 % lorsque l'on prend la classification à 99 catégories d'occupation.

<sup>28</sup> Seule différence statistiquement significative au seuil de 5 %.

**Tableau 2 : Les transitions entre les catégories d'occupation à Lima M., 1985-1990**

Annees	1990					Différences entre colonnes			
	1985	1	2	3	Total	Total colonnes	(1)-(2)	(1)-(3)	(2)-(3)
<b>Ensemble</b>									
	1	71,2	22,6	6,2	100	18,0	***	***	***
	2	11,5	75,2	13,3	100	47,3	***	NS	***
	3	3,9	34,2	61,9	100	34,7	***	***	***
<i>Total lignes</i>		19,6	51,5	28,9	100	100			
<b>Hommes</b>									
	1	74,3	21,0	4,8	100	21,2	***	***	***
	2	14,9	75,4	9,7	100	46,1	***	*	***
	3	4,3	27,8	67,9	100	32,7	***	***	***
<i>Total lignes</i>		24,0	48,3	27,7	100	100			
<b>Femmes</b>									
	1	63,4	26,8	9,8	100	13,0	***	***	**
	2	6,5	74,8	18,7	100	49,2	***	***	***
	3	3,4	42,9	53,8	100	37,8	***	***	*
<i>Total lignes</i>		12,7	56,5	30,8	100	100			
<b>Différences entre les sexes</b>									
		1	2	3					
	1	***	NS	NS					
	2	**	NS	**					
	3	NS	***	**					

Source : Panel LSMS 1985-1990, construction des auteurs.

Champ : Les individus de 26-64 ans en 1985 exerçant une activité en 1985 et 1990.

Note : Tests de différences entre hommes et femmes. \* différence significative au seuil de 10 %, \*\* de 5 %, \*\*\* de 1 %

Une analyse complémentaire par les « rapports de chances relatives » (*odds ratio*) permet de corriger les effets du changement de structure occupationnelle entre les deux dates<sup>29</sup>. Elle montre que la fluidité entre le groupe 2 et le groupe 3 (*odd ratio* = 10.24) est plus importante que celle entre le groupe 1 et le groupe 2 (*odd ratio* = 20.63).

### 2.3. Les flux entre les groupes occupationnels entre 1982 et 1990

Grâce à l'historique de l'emploi de 1985, il est possible d'observer l'ensemble de flux d'entrées et de sorties qui se produisent entre les groupes occupationnels en trois points du temps : 1982, 1985 et 1990 (cf. graphique 2). L'utilisation d'une période d'observation plus longue permet d'un peu mieux illustrer la complexité des flux qui se produisent entre les trois groupes occupationnels au cours du temps. Malheureusement, la faible taille de l'échantillon panel ainsi que le nombre important d'individus inactifs en 1982 (19 %) ne permet pas de faire une analyse par genre.

Par rapport aux analyses précédentes, on observe quelques résultats intéressants. En premier lieu, il apparaît que les transitions se produisant entre les groupes sont assez complexes. En deuxième lieu, le groupe 2 apparaît comme une sorte de « trappe » car c'est celui qui retient la proportion la plus élevée d'individus permanents, c'est-à-dire d'individus qui sont dans le même groupe en 1982, 1985 et

<sup>29</sup> Les rapports de chances relatives traduisent le résultat de la concurrence entre les individus de groupes occupationnels distincts pour atteindre des destinées différentes. Ils représentent l'inégalité relative entre deux individus, occupant en 1985 respectivement le groupe *i* et le groupe *i'*, pour atteindre le groupe *j'* plutôt que *j* (avec  $i' > i$  et  $j' > j$ ) en 1990. Il est défini, pour la génération *k* comme suit :

$$OR_{i-i', j-j'}^k = \frac{\frac{n_{ij}}{n_{ij'}}}{\frac{n_{i'j}}{n_{i'j'}}} = \frac{n_{ij}n_{i'j'}}{n_{i'j}n_{ij}}$$

où  $n_{ij}$  est le nombre d'observations dans la cellule (i, j) de la matrice de transition dont les

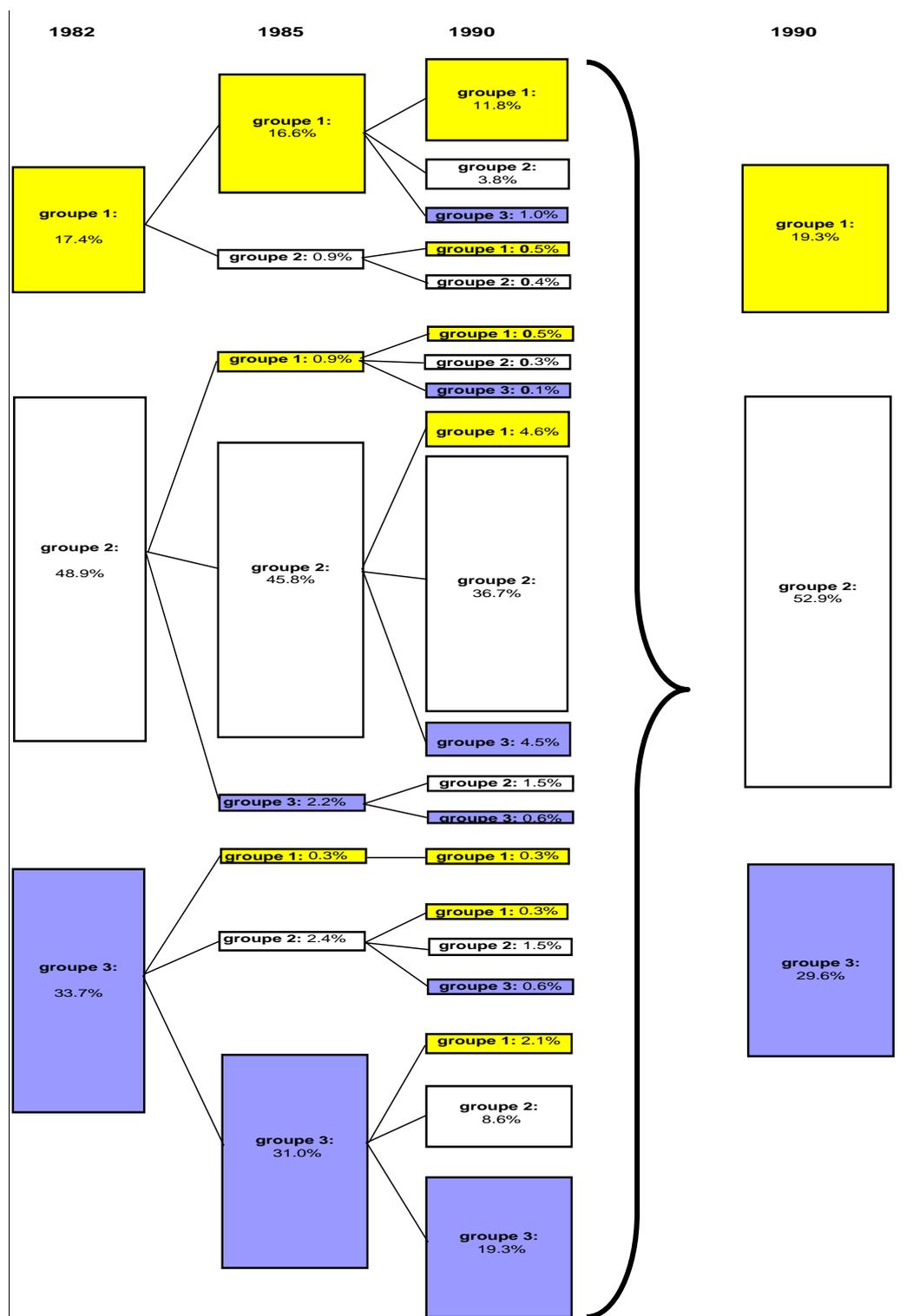
lignes *i* représentent les trois groupes d'occupations en 1985 et les colonnes *j* les trois groupes d'occupations en 1990.

La probabilité d'atteindre un groupe *j'* plutôt que *j* est  $OR_{i-i', j-j'}$  fois supérieure pour un individu du groupe *i'* qu'un autre du groupe *i*.

Si le *odd ratio* vaut 1, l'individu du groupe *i'* n'a pas d'avantage comparatif par rapport à celui du groupe *i* pour atteindre *j'*. La caractéristique des rapports de chances relatives est qu'ils donnent une mesure de l'association statistique entre deux variables indépendamment des distributions marginales.

1990<sup>30</sup> (75 % des personnes appartenant au groupe 2 en 1982 s’y trouvent en 1985 et 1990). Il est suivi de près par le groupe 1 (71 %) alors que la proportion d’individus permanents du groupe 3 n’est que de 56 %<sup>31</sup>. Enfin, comme dans les analyses précédentes, il apparaît que les transitions entre les groupes 1 et 3 sont insignifiantes.

**Graphique 2 : Transitions entre les groupes occupationnels entre 1982, 1985 et 1990 (%)**



Source : Panel LSMS 1985-1990 et informations sur l’historique de l’emploi de l’enquête LSMS 1985, construction des auteurs.  
 Champ : Les individus de 26-64 ans en 1985 exerçant une activité en 1982, 1985 et 1990.

<sup>30</sup> Cependant, c’est aussi le groupe le plus large, ce qui relativise ce résultat.

<sup>31</sup> Un test de différences montre que ces trois proportions, 75 %, 71 % et 56 % sont significativement différentes les unes des autres.

### 3. LES EFFETS DE L'ORIGINE SOCIALE SUR LA MOBILITE OCCUPATIONNELLE

On s'intéresse maintenant au rôle de l'origine sociale dans l'évolution de la carrière professionnelle.

#### 3.1. Le lien entre l'origine sociale et la mobilité occupationnelle par l'analyse descriptive

La réponse que l'on apportera quant à l'influence de l'origine sociale sur la mobilité occupationnelle peut dépendre de la manière dont on envisage la mobilité occupationnelle, si c'est de manière quantitative ou de manière qualitative. L'approche quantitative pense la mobilité en terme de mouvements. Une personne sera d'autant plus mobile qu'elle aura souvent changé d'occupation au cours d'une période donnée. En revanche, l'approche qualitative conçoit la mobilité occupationnelle en terme de trajectoires. Une personne sera d'autant plus mobile que son point de départ est éloigné de son point d'arrivée. Dans l'analyse descriptive, on traitera simultanément ces deux approches qui sont complémentaires.

##### 3.1.1 L'origine sociale influence-t-elle la mobilité occupationnelle en terme de mouvements ?

Une mesure simple de la mobilité occupationnelle en terme de mouvements est le nombre de changements d'occupation au cours d'une période déterminée. Grâce aux historiques d'emplois des enquêtes de 1985 et 1990, il a été possible de calculer, pour chaque individu du panel, le nombre de changements d'occupation<sup>32</sup> qu'il a vécu entre 1982 et 1990, c'est-à-dire durant huit ans de carrière. Si l'origine sociale conditionne la mobilité occupationnelle en termes de mouvements, alors la distribution du nombre de changements d'occupation est différente selon l'origine sociale.

L'origine sociale est ici mesurée uniquement par l'occupation du père. L'occupation des parents est codifiée dans l'enquête de 1985 en 9 catégories, assez peu pertinentes. De ces catégories ont pu être constitués trois groupes, celui des professions qualifiées (cadres, administrateurs, professions intellectuelles et libérales, employés administratifs), celui des agriculteurs et enfin un groupe qui est constitué du reste, c'est-à-dire des professions non qualifiées hors agriculture<sup>33</sup>. En analysant le niveau d'éducation moyen au sein de chacun des groupes, il apparaît que celui du groupe des professions qualifiées est de loin le plus élevé, suivi ensuite par celui du groupe des professions non qualifiées non agricoles, puis, loin derrière, par celui des professions agricoles.

Le tableau 3 montre que la distribution du nombre de changements d'occupation varie selon l'origine sociale, pour les hommes et pour les femmes pris séparément<sup>34</sup>. Les hommes d'origine la plus défavorisée sont significativement moins mobiles que ceux d'origine la plus favorisée. A l'inverse, les femmes dont le père exerçait une profession qualifiée (profession libérale, cadre, fonctionnaire ou employé administratif) connaissent significativement moins de changements professionnels que les autres femmes. Autrement dit, pour les femmes, une origine sociale élevée semble conférer une stabilité dans la carrière. Ces deux constatations laissent penser que l'origine sociale a un impact sur la mobilité occupationnelle lorsque celle-ci est vue en terme de mouvement. En est-il de même si la mobilité est vue en terme de trajectoire ?

---

<sup>32</sup> Le nombre de changements est mesuré entre les 99 groupes occupationnels afin d'obtenir l'analyse la plus fine possible de la mobilité.

<sup>33</sup> Comprend les activités de commerce, de transport, de construction, d'industrie, de services et des mines.

<sup>34</sup> Ce n'est pas le cas pour l'ensemble car les différences se compensent entre les genres.

**Tableau 3 : Nombre de changements d'occupation entre 1982 et 1990 selon l'origine sociale (%), Lima M.**

Occupation du père Nombre de changements	(1)	(2)	(3)	Ensemble	Significativité des différences			
	Profession qualifiée	Profession non qualifiée	Agriculteur		(1)-(2)	(1)-(3)	(2)-(3)	
Ensemble	0	46	39	50	45	NS	NS	NS
	1	47	46	37	43	NS	NS	NS
	2	7	13	11	11	NS	NS	NS
	3	1	2	2	1	NS	NS	NS
		100	100	100	100			
Homme	0	30	38	49	41	*	**	NS
	1	59	47	41	47	**	**	NS
	2	10	12	8	10	NS	NS	NS
	3	1	2	1	2	NS	NS	NS
		100	100	100	100			
Femme	0	73	43	51	52	*	*	NS
	1	26	43	30	34	NS	NS	NS
	2	1	14	17	13	NS	NS	NS
	3	0	0	2	1	NS	NS	NS
		100	100	100	100			

Source : Panel LSMS 1985-1990 et informations sur l'historique de l'emploi de la LSMS 1985, construction des auteurs.

Note : NS : différence non significative, \* significative au seuil de 10 %, \*\* de 5 %.

Champ : Les individus de 26-64 ans en 1985 exerçant une activité en 1982, 1985 et 1990.

### 3.1.2 L'origine sociale influence-t-elle la mobilité occupationnelle en terme de trajectoires ?

Un fils d'agriculteur qui était technicien en 1982, a-t-il la même probabilité d'être cadre en 1990 qu'un fils de cadre, lui-même technicien en 1982 ? L'examen des matrices de transitions montre que non. Les probabilités d'atteindre une certaine position en 1990 étant donnée sa position en 1982 diffèrent selon l'origine sociale (cf. tableau 4).

Tout d'abord, les chances de mobilité ascendante ne sont pas les mêmes pour tous : la probabilité de connaître une mobilité ascendante vers une profession qualifiée (groupe 1) est significativement plus élevée pour les personnes dont le père exerçait lui-même une profession qualifiée que pour les autres<sup>35</sup>. Il existe donc au Pérou un phénomène de « contre-mobilité » tel que le définissent les sociologues : les personnes d'origine sociale favorisée, même si elles exercent en début de carrière une occupation qui leur procure un statut moins élevé que celui de leurs parents vont connaître une plus forte mobilité ascendante, soit de par leur plus forte motivation, soit grâce à leur capital social hérité, et recouvrer ainsi leur statut d'origine.

Ensuite, la probabilité de connaître une mobilité descendante vers les professions non-qualifiées est significativement plus faible pour ceux d'origine sociale élevée que pour les autres<sup>36</sup>.

En revanche, il n'y a pas de différence significative entre personnes d'origine sociale distinctes quant à la probabilité d'être immobile. Ceci pourrait s'expliquer, tout comme dans le tableau 3 précédent, par des comportements différents et compensatoires entre hommes et femmes.

<sup>35</sup> 18 % de ceux dont le père exerçait une profession qualifiée connaissent une mobilité ascendante entre 1982 et 1990 du groupe 2 vers le groupe 1. Cette proportion n'est que de 10 % pour ceux dont le père avait une profession non qualifiée hors agriculture et 7 % pour ceux dont le père était agriculteur. La différence entre cette première proportion et les deux autres est significative au seuil de 10 %. Dans le cas d'une mobilité ascendante du groupe 3 vers le groupe 1, ces proportions sont respectivement 26 %, 14 % et 3 % et les différences entre la première proportion et les deux autres sont significatives au seuil de 1 %.

<sup>36</sup> La probabilité de passer du groupe 1 au groupe 3 entre 1982 et 1990 est nulle pour ceux dont le père exerçait une profession qualifiée alors qu'elle est de 6 % pour ceux dont le père avait une profession non qualifiée hors agriculture et 16 % pour ceux dont le père était agriculteur, les différences entre ces trois proportions étant significatives au seuil de 5 et 10 %. De plus, ceux d'origine la plus favorisée ont une probabilité significativement plus faible de passer du groupe 2 au groupe 3 que ceux d'origine intermédiaire (0,01 vs. 0,14).

**Tableau 4 : Matrices de transition entre 1982 et 1990 en fonction de l'origine sociale**

Occupation du père	Occupation en 1990	Groupe 1	Groupe 2	Groupe 3	Ensemble	<i>N obs</i>	
<b>(1)</b> Profession qualifiée	Groupe 1	78	22	0	100	36	
	Groupe 2	18	81	1	100	42	
	Groupe 3	26	22	52	100	11	
<b>(2)</b> Profession non qualifiée hors agri.	Groupe 1	65	27	7	100	55	
	Groupe 2	10	76	14	100	155	
	Groupe 3	14	31	54	100	82	
<b>(3)</b> Agriculteur	Groupe 1	66	17	16	100	18	
	Groupe 2	7	84	9	100	109	
	Groupe 3	3	28	69	100	120	
						<b>628</b>	
<b>Significativité des différences</b> Occupation en 1982	Occupation en 1990	Groupe 1	Groupe 2	Groupe 3			
	<b>Différences</b>						
		(1)-(2)	NS	NS	*		
	Groupe 1	(1)-(3)	NS	NS	**		
		(2)-(3)	NS	NS	NS		
		(1)-(2)	*	NS	**		
	Groupe 2	(1)-(3)	*	NS	NS		
		(2)-(3)	NS	NS	NS		
		(1)-(2)	***	NS	NS		
	Groupe 3	(1)-(3)	***	NS	NS		
		(2)-(3)	NS	NS	NS		

Source : Panel LSMS 1985-1990 et informations sur l'historique de l'emploi de l'enquête LSMS 1985, construction des auteurs. Note : NS : différence non significative, \* significative au seuil de 10%, \*\* de 5%, \*\*\* au seuil de 10%.  
Champ : Les individus de 26-64 ans en 1985 exerçant une activité en 1982, 1985 et 1990.

### 3.2. Le lien entre l'origine sociale et la mobilité occupationnelle par l'analyse économétrique

#### 3.2.1 Test de l'hypothèse de persistance de l'effet de l'origine sociale au cours de la carrière par les modèles log-linéaires

L'analyse précédente à partir des matrices de transitions ne prend pas en compte l'évolution de la structure occupationnelle entre 1985 et 1990. Autrement dit, si l'on a constaté que l'origine a un effet significatif sur la mobilité occupationnelle observée, cette mobilité occupationnelle est entachée de la mobilité structurelle. Les modèles log-linéaires (cf. annexe 3 pour une présentation détaillée de ces modèles,) permettent de tester d'une part si l'origine sociale a un effet sur l'occupation en 1990 et ceci, quelle que soit l'évolution de la structure occupationnelle entre 1985 et 1990, c'est-à-dire en neutralisant l'évolution des distributions marginales des occupations<sup>37</sup>. Ils permettent d'autre part de tester l'effet de l'origine sociale sur l'occupation en 1990 en contrôlant par l'occupation en 1985. Autrement dit, ils testent l'hypothèse selon laquelle il existe une association statistique entre l'occupation en 1990 et l'origine sociale, conditionnellement à l'occupation en 1985 et ceci en contrôlant par la mobilité structurelle entre 1985 et 1990.

<sup>37</sup> La structure occupationnelle étant ici très simplifiée puisqu'elle se réduit aux trois groupes occupationnels définis antérieurement.

### 3.2.1.1 Les modèles

Trois modèles log-linéaires sont testés, chacun postulant une hypothèse différente. Le premier modèle suppose que l'occupation est indépendante de l'origine sociale. Ce modèle implique une interaction entre l'occupation en 1985 et celle en 1990 mais aucun lien entre l'origine et l'occupation en 1985 ou celle en 1990. Ce modèle teste donc si l'occupation en 1990 est corrélée à celle de 1985, quelle que soit l'évolution de la structure occupationnelle entre les deux dates :

$$\ln(F_{i,j,k}) = \mu + \lambda_i^O + \lambda_j^{D^{85}} + \lambda_k^{D^{90}} + \lambda_{j,k}^{D^{85}D^{90}} \quad (1)$$

où  $O$  désigne l'origine sociale, approchée par l'occupation principale exercée par le père au cours de sa vie (3 modalités),  $D^{85}$  l'occupation principale exercée par l'enquêté en 1985 (3 modalités),  $D^{90}$  l'occupation principale exercée en 1990 (3 modalités),  $F_{i,j,k}$  la fréquence de la cellule  $(i, j, k)$  de la table à trois entrées  $(O, D^{85}, D^{90})$  ;  $\lambda_i^O, \lambda_j^{D^{85}}, \lambda_k^{D^{90}}$  sont les effets principaux sur la distribution des individus de  $O, D^{85}, D^{90}$  respectivement et enfin,  $\lambda_{j,k}^{D^{85}D^{90}}$  est l'interaction partielle entre  $D^{85}$  et  $D^{90}$ , étant donnée  $O$ . Ce modèle est peu réaliste mais joue le rôle de modèle de référence.

Le second modèle suppose que seule l'occupation en 1985 dépend de l'origine sociale. En plus de prendre en compte l'interaction entre  $D^{85}$  et  $D^{90}$ , il introduit une interaction partielle entre l'origine  $O$  et  $D^{85}$ , étant donnée l'occupation en 1990,  $\lambda_{i,j}^{OD^{85}}$ . Le modèle (2) postule donc que les matrices de transition sont différentes pour chaque groupe d'origine sociale uniquement de par l'effet de l'origine sociale sur l'occupation en 1985 :

$$\ln(F_{i,j,k}) = \mu + \lambda_i^O + \lambda_j^{D^{85}} + \lambda_k^{D^{90}} + \lambda_{j,k}^{D^{85}D^{90}} + \lambda_{i,j}^{OD^{85}} \quad (2)$$

Le troisième modèle teste l'hypothèse selon laquelle à la fois les occupations en 1985 et celles en 1990 dépendent de l'origine sociale. Dans ce modèle, on prend donc en compte une interaction supplémentaire, celle entre l'origine  $O$  et  $D^{90}$ , étant donnée l'occupation en 1985,  $\lambda_{i,k}^{OD^{90}}$ . Ce modèle suppose donc que l'origine sociale a un rôle au cours de la carrière dans la mesure où il n'est pas possible de reproduire les matrices de transitions en considérant qu'il y a indépendance de l'occupation en 1990 à l'origine sociale, conditionnellement à l'occupation en 1985. Ceci signifie que si l'on impose par exemple aux entreprises de remplacer en 1985 tous les travailleurs d'une certaine origine sociale par ceux d'une autre origine sociale, on n'obtiendra pas la même matrice de transitions entre 1985 et 1990 que celle que l'on aurait obtenue sans cette permutation. Le modèle 3 est défini de la façon suivante :

$$\ln(F_{i,j,k}) = \mu + \lambda_i^O + \lambda_j^{D^{85}} + \lambda_k^{D^{90}} + \lambda_{j,k}^{D^{85}D^{90}} + \lambda_{i,j}^{OD^{85}} + \lambda_{i,k}^{OD^{90}} \quad (3)$$

La comparaison de ces trois modèles permet de valider ou d'invalider l'hypothèse de l'existence d'une interaction partielle entre l'occupation en 1990 et l'origine sociale, étant donnée l'occupation en 1985. Cette hypothèse est retenue si le troisième modèle est celui qui reconstitue le mieux les matrices de transition, avec la plus grande parcimonie, c'est-à-dire celui qui s'ajuste le mieux aux données observées tout en contenant le moins de paramètres. Plusieurs indicateurs nous permettent d'évaluer ces trois modèles. Ils sont présentés en annexe 4 et le tableau 5 donne leur valeur.

**Tableau 5 : Modèles log-linéaires avec l'origine sociale, l'occupation en 1985, l'occupation en 1990**

Modèles	ddl	$G^2$	P	$rG^2$	$\Delta\%$	BIC	$\chi^2$
(1) L'activité professionnelle est indépendante de l'origine sociale (O)(D <sup>85</sup> D <sup>90</sup> )	16	120,7	0,000	0,0	11,5	14,3	
(2) L'activité en 1985 dépend de l'origine sociale (O)(D <sup>85</sup> D <sup>90</sup> )(OD <sup>85</sup> )	12	31,6	0,002	73,8	5,5	-48,2	
(3) L'activité en 1985 et en 1990 dépendent de l'origine sociale (O)(D <sup>85</sup> D <sup>90</sup> )(OD <sup>85</sup> )(OD <sup>90</sup> )	8	10,2	0,252	91,6	3,2	-43,0	
Test de l'effet de X: Modèle (3)-(Modèle(3)-X)							
X=(OD <sup>90</sup> )	4		0,000				21,4
X=(OD <sup>85</sup> )	4		0,000				22,79

Source : Panel LSMS 1985-1990, construction des auteurs

Champ : Les individus de 26-64 ans en 1985 et plus exerçant une activité en 1985 et en 1990

### 3.2.1.2 Les résultats

Les modèles (1) et (2) ne reproduisent pas correctement les données puisque le test construit à partir de la statistique de vraisemblance  $G^2$  conclue pour les deux modèles que les données estimées sont significativement différentes des données empiriques<sup>38</sup>. Les hypothèses selon lesquelles l'activité professionnelle est indépendante de l'origine sociale ou que seule l'activité en 1985 est corrélée à l'origine sociale doivent être rejetées<sup>39</sup>. Le modèle (3) en revanche s'ajuste correctement aux données. D'après l'indice de dissimilarité, moins de 4 % de l'effectif de la matrice estimée par ce modèle doit être déplacé pour obtenir la matrice observée. De plus, lorsque l'on introduit un effet de l'origine sociale sur l'occupation en 1990, étant donnée l'occupation en 1985 –modèles (3) –, plus de 90 % des observations qui sont mal reproduites sous le modèle d'indépendance –modèle (1)- sont maintenant correctement reproduites<sup>40</sup>. Cette proportion n'est que de 74 % si on prend en compte uniquement un effet de l'origine sociale sur l'occupation en 1985 –modèle (2)-.

Ainsi, le modèle (3) doit être préféré au modèle (2), ce qui signifie que l'influence de l'origine sociale n'est pas entièrement captée par son effet sur l'occupation en 1985. Un test au sein du modèle (3) confirme ce résultat en montrant que l'effet de l'origine sociale sur l'occupation en 1990, contrôlé par l'occupation en 1985,  $\lambda_{i,k}^{OD_{90}}$  est significatif. Cet effet est cependant de la même ampleur<sup>41</sup> que celui de l'origine sociale sur l'occupation en 1985,  $\lambda_{i,j}^{OD_{85}}$ .

Cette analyse a donc permis de mettre en évidence que l'origine sociale a une influence sur l'occupation en 1990 conditionnellement à l'occupation en 1985 et que son influence ne semble pas cumulative au cours du temps. Cependant, il n'est pas possible de conclure de manière robuste que l'effet de l'origine sociale influence de manière persistante l'évolution de la carrière professionnelle car pour cela, il faudrait prendre en compte les relations de causalité entre l'origine sociale, l'occupation en 1985 et celle en 1990, ce que ne permettent pas les modèles log-linéaires.

<sup>38</sup>  $P < 0,01$  ce qui signifie que l'on rejette au seuil de 1% l'hypothèse selon laquelle les données estimées par ces modèles ne sont pas significativement différentes des données empiriques.

<sup>39</sup> On peut douter de la validité du test construit à partir de la statistique de vraisemblance pour rejeter le modèle 2 car il dépend de la taille de l'échantillon (qui n'est cependant pas élevée ici). Cependant le rejet du modèle 2 est confirmée par le test de Chi-2 puisque la valeur du quantile à 95 % d'un Chi-2 à 4 degrés de liberté est égale à 9,5 et la différence entre les statistiques de vraisemblance des 2 modèles à 21,4 (cf. annexe 4).

<sup>40</sup> D'après la statistique  $rG^2$ .

<sup>41</sup> La statistique de  $\chi^2$  est à peu près égale à 22 pour les deux effets.

### 3.2.2 Test de l'effet de l'origine sociale sur la carrière professionnelle par le modèle probit bivarié

L'analyse économétrique par les modèles log-linéaires conclue à l'existence d'un lien entre la mobilité occupationnelle et l'origine sociale. Cependant, il est nécessaire de confirmer de façon plus robuste l'existence de ce lien en vérifiant si l'effet de l'origine sociale persiste au cours du temps par un modèle qui prenne en compte les relations de causalité. Ce modèle doit en outre autoriser la prise en compte de plusieurs dimensions de l'origine sociale, et plus seulement l'occupation du père, et permettre l'étude des effets de l'origine sociale « toutes choses étant égales par ailleurs », c'est-à-dire en contrôlant les effets des caractéristiques individuelles<sup>42</sup>.

On cherche à savoir comment l'origine sociale sous ses différentes dimensions agit sur l'occupation en 1990, indépendamment du rôle qu'elle a déjà eu jusqu'en 1985. On suppose alors que la position en 1985 contient tous les effets qu'a déjà eu l'origine sociale sur le parcours d'une personne : l'effet de l'environnement familial au moment de la formation, au moment du premier emploi, font, avec d'autres facteurs, que la personne occupe telle place en 1985 plutôt qu'une autre. Tester économétriquement la persistance de l'origine sociale au cours de la carrière professionnelle consiste alors à vérifier si l'origine sociale explique de manière significative la position occupationnelle atteinte en 1990 une fois que l'on contrôle par la position en 1985. Contrôler par la position permet de capter d'une part la dynamique occupationnelle entre 1985 et 1990 et d'autre part, l'effet indirect de l'origine sociale, c'est-à-dire l'effet qu'a eu l'origine sociale jusqu'en 1985.

#### 3.2.2.1 Le choix du modèle approprié

Il n'est pas facile de trouver un modèle économétrique qui soit vraiment adapté à notre problématique et qui puisse être estimé avec les données dont on dispose. La variable dépendante du modèle, l'occupation en 1990, est dichotomique. Elle prend la valeur 1 si l'individu appartient à un certain groupe occupationnel et 0 sinon. Le modèle économétrique retenu est donc non linéaire.

On suppose dans un premier temps que le modèle adapté est un modèle de type probit simple :

$$D^{90*} = \alpha_1 X + \beta_1 O + \gamma_1 D^{85} + u_1, \text{ où } \begin{cases} D^{90} = 1 & \text{si } D^{90*} > 0 \\ D^{90} = 0 & \text{si } D^{90*} \leq 0 \end{cases} \quad (4)$$

avec  $\alpha_1$ ,  $\beta_1$ ,  $\gamma_1$  des vecteurs de paramètres,  $D^{85}$  et  $D^{90}$  les variables dichotomiques observées du groupe d'occupation en 1985 et en 1990,  $D^{90*}$  une variable latente non observée,  $X$  des variables de caractéristiques individuelles,  $O$  des variables d'origine sociale et  $u_1$  le vecteur des variables inobservées.

Cependant, il est clair que la position occupée en 1985,  $D^{85}$ , dépend elle aussi de l'origine sociale :

$$D^{85*} = \alpha_2 X + \beta_2 O + u_2, \text{ où } \begin{cases} D^{85} = 1 & \text{si } D^{85*} > 0 \\ D^{85} = 0 & \text{si } D^{85*} \leq 0 \end{cases} \quad (5)$$

<sup>42</sup> A cause de l'effectif trop faible de notre échantillon, il n'était en effet pas possible d'introduire dans l'analyse log-linéaire plus de trois variables. Le seul contrôle qui a pu être fait est celui de l'occupation en 1985. Goux et Maurin (1997) contrôle également dans leur analyse par le niveau d'éducation mais ils ont à leur disposition un échantillon beaucoup plus important que le notre.

avec  $\alpha_2, \beta_2, \gamma_2$  des vecteurs de paramètres,  $X$  des variables de caractéristiques individuelles,  $D^{85*}$  une variable latente non observée et  $u_2$  le vecteur des résidus.

L'estimation du modèle (4) comporte alors un biais dans la mesure où  $D^{85}$  est endogène au modèle. En effet, il est probable que dans le vecteur des résidus  $u_1$  se trouvent des composantes inobservées de l'origine sociale non captées par le vecteur  $O$  expliquant la position occupationnelle en 1990. On peut penser par exemple au capital social des parents qui n'est qu'imparfaitement pris en compte par les variables  $O$ , ce dernier pouvant jouer un rôle significatif sur la détermination d'une occupation. Si cette variable est incluse dans  $u_1$ , elle l'est également dans  $u_2$ . Dans ce cas,  $u_1$  et  $D^{85}$  ne sont plus orthogonales :

$$\text{corr}(u_1, D^{85}) \neq 0 \text{ ou encore } \text{corr}(u_1, u_2) = \rho.$$

Pour prendre en compte ce biais d'endogénéité, il convient d'estimer simultanément les équations (4) et (5), grâce à un modèle probit bivarié. Ce type de modèle permet en effet de contrôler l'endogénéité qui existe entre deux choix qui sont liés car il prend en compte la corrélation entre les termes d'erreurs. Le modèle à estimer ici comprend cependant une spécificité par rapport à un modèle probit bivarié classique : l'occupation en 1985 est une variable explicative dans l'équation (4) et dans le même temps, la variable latente correspondant à l'occupation en 1985 est la variable expliquée de l'équation (5). Maddala (1983, p.122-123) montre que les paramètres de ce type de modèle probit bivarié sont identifiables à la condition que l'équation (5) comprenne une variable qui ne soit pas incluse dans l'équation (4). Pour corriger le biais d'endogénéité, il faut donc estimer les équations (4) et (5) simultanément par un modèle probit bivarié mais en incluant dans l'équation (5) un instrument  $I$ , c'est-à-dire une variable qui explique la position occupationnelle en 1985 tout en étant indépendante de la position en 1990. Le modèle à estimer est alors :

$$\left\{ \begin{array}{l} D^{90*} = \alpha_1 X + \beta_1 O + \gamma_1 D^{85} + u_1 \\ D^{85*} = \alpha_2 X + \beta_2 O + \gamma_2 I + u_2 \end{array} \right. , \text{ où } \left\{ \begin{array}{ll} D^{90} = 1 & \text{si } D^{90*} > 0 \\ D^{90} = 0 & \text{si } D^{90*} \leq 0 \\ D^{85} = 1 & \text{si } D^{85*} > 0 \\ D^{85} = 0 & \text{si } D^{85*} \leq 0 \end{array} \right. \quad (6)$$

$$\text{Avec } \left\{ \begin{array}{l} E(u_1) = E(u_2) = 0 \\ \text{Var}(u_1) = \text{Var}(u_2) = 1 \text{ et } (u_1, u_2) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \rho & 1 \\ 1 & \rho \end{pmatrix}\right) \\ \text{Corr}(u_1, u_2) = \rho \end{array} \right.$$

On contraint les termes d'erreurs à avoir une variance unitaire pour pouvoir identifier le modèle, les variables  $D^{85}$  et  $D^{90}$  étant dichotomiques.

La fonction de vraisemblance à maximiser de ce modèle s'écrit :

$$L_i = \prod P_{11}^{D^{90}D^{85}} P_{10}^{D^{90}(1-D^{85})} P_{01}^{(1-D^{90})D^{85}} P_{00}^{(1-D^{90})(1-D^{85})}$$

où  $P_{11} = \Pr(D^{90} = 1, D^{85} = 1) = \Phi_2(\alpha_1 X + \beta_1 O + \gamma_1, \beta_2 O + \gamma_2 I, \rho)$ ,  
 $P_{10} = \Pr(D^{90} = 1, D^{85} = 0) = \Phi_2(\alpha_1 X + \beta_1 O, -\beta_2 O - \gamma_2 I, -\rho)$ ,  
 $P_{01} = \Pr(D^{90} = 0, D^{85} = 1) = \Phi_2(-\alpha_1 X - \beta_1 O - \gamma_1, \beta_2 O + \gamma_2 I, -\rho)$  et  
 $P_{00} = \Pr(D^{90} = 0, D^{85} = 0) = \Phi_2(-\alpha_1 X - \beta_1 O, -\beta_2 O - \gamma_2 I, \rho)$ , et  
 $\Phi_2$  est la fonction de répartition de la loi normale bivariée.

### 3.2.2.2 Le choix d'un instrument

Un premier instrument auquel on peut penser est la position occupationnelle de l'individu à une date antérieure à 1985. Dans l'enquête, grâce aux historiques d'emploi, il est possible de reconstituer cette position pour l'année 1982. Cet instrument est valable mais sous une hypothèse très forte, à savoir qu'il n'y ait pas d'effet fixe dans les résidus  $u_1$  et  $u_2$ . En effet, le biais d'endogénéité n'est corrigé que si toutes les composantes des résidus n'influencent plus la trajectoire occupationnelle entre 1985 et 1990, ie. si ces dernières ne jouent qu'à un moment ponctuel de la carrière, antérieur ou égal à 1985, notamment lors du premier emploi. Or il est probable que certaines composantes des résidus interfèrent sur la trajectoire occupationnelle de façon persistante. C'est le cas de caractéristiques individuelles inobservées telles que le talent : toutes choses égales par ailleurs, une personne talentueuse aura *a priori* une probabilité plus forte de connaître une trajectoire occupationnelle ascendante qu'une personne dépourvue de talent. Ainsi, en présence d'effets fixes individuels, la position occupationnelle en 1982 n'est pas un bon instrument car elle est elle-même corrélée aux effets fixes.

Un second instrument permettrait de lever cette hypothèse trop forte de la non-existence d'effet fixe : la position occupationnelle du conjoint en 1982<sup>43</sup>. La position occupationnelle du conjoint est un bon instrument pour deux raisons. D'une part elle est susceptible d'influer la position occupationnelle de l'individu à un moment donné –antérieur ou égal à 1985– par le canal suivant : le conjoint peut faire bénéficier l'individu de son réseau social, en particulier lors de l'entrée de ce dernier sur le marché du travail, ce réseau étant approché par la position qu'occupe le conjoint en 1982. D'autre part, le conjoint a des caractéristiques individuelles différentes de celles de l'individu, notamment en terme de talent ou de motivation. Cet instrument n'est donc pas *a priori* corrélé aux effets fixes. Cependant, ceci n'est vrai que sous certaines hypothèses. La première est que le « marché du mariage » intervienne avant l'entrée de l'individu sur le marché du travail ou bien qu'il n'y ait pas d'« assortative mating », c'est-à-dire que les individus ne se marient pas avec des personnes ayant des caractéristiques similaires aux leurs, telles que le talent. La seconde est que le capital social transféré n'intervienne qu'au début de la carrière, mais pas dans son déroulement. Il est possible de contrôler en partie la première hypothèse en ne considérant que les personnes qui se sont mariées avant 1982. On ne connaît pas la date du mariage mais l'âge du premier enfant. En ne gardant que les personnes qui ont un conjoint et au moins un enfant âgé de plus de huit ans, il est possible d'effectuer ce contrôle. Une limite évidente de ce test est que l'on crée un très fort biais de sélection, notamment si les comportements de fertilité sont différenciés selon les origines sociales et les trajectoires occupationnelles.

<sup>43</sup> Ce choix d'instrument s'inspire de Lam et Schoeni (1993) qui utilisent des variables d'origine sociale du conjoint comme instrument pour estimer les effets de l'origine sociale sur les rémunérations des hommes au Brésil.

### 3.2.2.3 L'estimation du modèle

Dans l'équation (4) la variable dépendante ( $D^{90*}$ ) est expliquée par trois types de variables :  $X$ ,  $O$  et  $D^{85}$ . Les caractéristiques individuelles  $X$  retenues sont le sexe de l'individu, son âge et une variable composite de son sexe et son âge. Les variables d'origine sociale  $O$  dont nous disposons ici sont : le niveau d'éducation du parent le plus éduqué (sans éducation, primaire, secondaire et supérieur), l'occupation du père (profession libérale/cadre/employé de l'administration, agriculteur, autre), si la personne est née ou non dans une ville et enfin, si elle a achevé son éducation dans une école publique plutôt que privée. Cette dernière variable est une proxy du niveau de vie des parents puisque, au Pérou, les écoles privées sont souvent onéreuses mais bénéficient d'une bien meilleure réputation que les écoles publiques. Les parents dotés de moyens financiers suffisants préféreront donc investir dans une école privée pour leurs enfants. Enfin, le troisième type de variables est le groupe occupationnel de la personne en 1985 ( $D^{85}$ ) permettant de capter la dynamique occupationnelle entre 1985 et 1990 et l'effet qu'a eu l'origine sociale jusqu'en 1985.

Pour l'estimation nous avons procédé de façon hiérarchique c'est-à-dire que nous avons estimé dans un premier temps les modèles 1 et 1bis dans lesquels les deux variables endogènes  $D^{85}$  et  $D^{90}$  sont appartenir ou non au groupe d'occupation le plus favorisé (groupe 1) versus appartenir aux deux autres groupes d'occupation. Dans un second temps, ont été estimés les modèles 2 et 2bis dont les variables endogènes sont appartenir au groupe d'occupation 2 versus appartenir au groupe 3<sup>44</sup>.

Les modèles 1 et 2 supposent l'absence d'effet fixe individuel dans les résidus et instrumentent donc le groupe d'occupation en 1985 par celui en 1982. Les modèles 1bis et 2bis acceptent l'existence d'effets fixes individuels mais ne sont justifiés que si le capital social du conjoint transféré n'intervient que ponctuellement et non pas au cours de la carrière, entre 1985 et 1990. Les estimations à partir de ces deux derniers modèles sont de plus réalisées à partir d'un échantillon relativement faible puisqu'il ne comprend que les personnes vivant en concubinage qui ont au moins un enfant de plus de 8 ans et elles souffrent *a priori* d'un biais de sélection. Nous ne présentons ici que les résultats des estimations des modèles 1bis et 2bis<sup>45</sup>. Ces résultats sont présentés dans le tableau 6.

---

<sup>44</sup> Une autre stratégie aurait été d'estimer un modèle probit bivarié ordonné mais celle-ci a été écartée pour des contraintes techniques.

<sup>45</sup> Les estimations des modèles 1 et 2 se trouvent dans l'annexe 5.

**Tableau 6 : Résultats des estimations à partir des modèles probit bivariés**

	Modèle 1bis		Modèle 2bis	
	Coefficients	Sign.	Coefficients	Sign.
<b>Equation 1: Le groupe occupationnel en 1990</b>				
<i>Variable dépendante</i>	1=Être du groupe 1 en 1990 0=Être du groupe 2 ou 3		1=Être du groupe 2 en 1990 0=Être du groupe 3	
<b>Variables sur les caractéristiques individuelles</b>				
Sexe (femmes = 1)	0,36		0,17	
Age	0,01		0,00	
<i>Sexe*Age (réfèrent: avoir plus de 40 ans)</i>				
Femme de moins de 40 ans	-0,66		0,19	
Homme de moins de 40 ans	0,35		0,10	
<b>Variables d'origine sociale</b>				
<i>Niveau d'éducation des parents (réfèrent: Education primaire)</i>				
Sans éducation	-0,88 **		-0,32	
Education secondaire ou supérieure	0,00		-0,32	
<i>Activité du père (réfèrent: Agriculteur)</i>				
Professionnel ou employé dans l'administration	0,11		-0,49	
Autres hors l'agriculture	0,13		-0,32	
<i>Lieu de naissance (ville=1)</i>	0,00		-0,05	
<i>Type d'école (publique = 1)</i>	0,11		-0,73 **	
<b>Le groupe occupationnel en 1985</b>				
Être du groupe 1 (réfèrent: être du groupe 2 ou 3)	2,67 ***			
Être du groupe 2 (réfèrent: être du groupe 3)			1,96 ***	
<b>Constante</b>	-2,14 **		-0,04	
<b>Equation 2: Le groupe occupationnel en 1985</b>				
<i>Variable dépendante</i>	1=Être du groupe 1 en 1985 0=Être du groupe 2 ou 3		1=Être du groupe 2 en 1985 0=Être du groupe 3	
<b>Variables sur les caractéristiques individuelles</b>				
Sexe (femmes = 1)	-0,86 ***		-0,15	
Age	-0,05 **		0,00	
<i>Sexe*Age (réfèrent: avoir plus de 40 ans)</i>				
Femme de moins de 40 ans	-0,37		-0,28	
Homme de moins de 40 ans	-1,31 ***		-0,22	
<b>Variables d'origine sociale</b>				
<i>Niveau d'éducation des parents (réfèrent: Education primaire)</i>				
Sans éducation	-0,35		-0,05	
Education secondaire ou supérieure	0,16		0,59 **	
<i>Activité du père (réfèrent: Agriculteur)</i>				
Professionnel ou employé dans l'administration	1,06		0,86 **	
Autres hors l'agriculture	0,18		0,15	
<i>Lieu de naissance (ville=1)</i>	-0,02		0,06	
<i>Type d'école (publique = 1)</i>	0,02		0,26	
<b>Instrument</b>				
<i>Le groupe occupationnel du conjoint en 1982 (réfèrent: inactivité)</i>				
Être du groupe 1	1,25 ***		1,72 ***	
Être du groupe 2	0,10		0,56 ***	
Être du groupe 3	-0,16		-0,07	
<b>Constante</b>	1,29		-0,48	
<b>Rho</b>	-0,77		-0,56	
<b>Log Likelihood</b>	-197,57		-286,57	
<b>N. observations</b>	318		247	

Source : Panel LSMS 1985-1990, construction des auteurs.

Champ : Les individus de 26-64 ans en 1985 et plus exerçant une activité en 1985 et en 1990.

Note : \* coefficient significatif au seuil de 10 %, \*\* au seuil de 5 %, \*\*\* au seuil de 1 %.

Dans chacun de ces deux modèles, une variable d'origine sociale a un effet significatif sur la probabilité d'atteindre un certain groupe occupationnel en 1990, étant donnée sa position en 1985. Dans le modèle 1bis, l'éducation du parent le plus éduqué a un effet significatif sur la probabilité de rester dans le groupe le plus favorisé (groupe 1) si on y était déjà en 1985 ou de réussir à y rentrer sinon. En effet, ceux dont les parents n'ont pas achevé leurs études primaires sont dotés d'une probabilité plus faible que les autres. Dans le modèle 2bis, le fait d'avoir achevé ses études dans un établissement privé augmente la probabilité de connaître une mobilité ascendante entre 1985 et 1990 du groupe 3 vers le groupe 2 ou d'éviter une mobilité descendante vers le groupe 3 si la personne était dans le groupe 2 en 1985.

Pour les deux modèles, le groupe occupationnel du conjoint a bien un effet significatif sur le groupe occupationnel atteint en 1985 et cet effet joue dans le sens attendu. De plus, l'effet de l'occupation en 1985 est bien positif et très significatif.

L'estimation de ces modèles ne nous permet pas de confirmer ou d'infirmer l'hypothèse selon laquelle l'effet de l'origine sociale persiste au cours de la trajectoire professionnelle car elle repose sur des hypothèses très fortes. Cependant, elle participe à un faisceau de présomption qui suggère plutôt qu'une telle persistance a lieu. Il semble que pour atteindre ou se maintenir dans les professions qualifiées ce soit surtout le capital culturel de la famille qui influe sur la trajectoire professionnelle. En revanche, pour connaître une mobilité ascendante vers les professions intermédiaires lorsque l'on exerçait une occupation non qualifiée en 1985 ou pour ne pas « tomber » vers une profession non qualifiée si l'on vient du groupe 2, il semble que le capital économique des parents lors de l'éducation des individus, approché par le type d'école fréquenté, soit déterminant.

## CONCLUSION

Le premier objectif de cet article était de caractériser la mobilité occupationnelle à Lima. Malgré le manque de finesse de la classification en trois groupes s'expliquant par la faible taille de l'échantillon, nous montrons l'importance et la complexité de cette mobilité. Cette mobilité est d'autant plus importante qu'elle concerne les individus en bas de l'échelle occupationnelle, *i.e.* entre les deux groupes occupationnels les moins favorisés. Les femmes semblent se distinguer des hommes en étant plus mobiles mais avec des trajectoires occupationnelles de plus faible ampleur.

Le second objectif était de tester l'hypothèse selon laquelle l'effet de l'origine sociale ne s'atténue pas avec le temps mais, au contraire, persiste tout au long de la carrière professionnelle. Les résultats de l'analyse descriptive de l'effet de l'origine sociale, mesurée par l'activité du père, sur la mobilité occupationnelle vont dans le sens de cette hypothèse. Ils font notamment apparaître un phénomène de contre-mobilité pour les personnes d'origine sociale élevée, c'est-à-dire que ces dernières ont plus de chances de connaître une mobilité occupationnelle ascendante que les autres. Cette analyse est complétée par l'estimation de modèles log-linéaires. Elle montre que l'origine sociale a une influence sur l'occupation en 1990 conditionnellement à l'occupation en 1985. Cependant, ces deux types d'analyse ne peuvent être considérés comme un test rigoureux de l'hypothèse de persistance du rôle de l'origine sociale sur la mobilité occupationnelle, car elles ne permettent ni de contrôler par les caractéristiques individuelles, ni d'introduire des relations de causalité. C'est pourquoi une analyse économétrique d'un autre type a été menée, par l'estimation d'un modèle probit bivarié. Les résultats de cette estimation vont dans le sens des analyses précédentes. Cependant, cette analyse se heurte à des difficultés méthodologiques –à savoir le traitement de l'endogénéité des variables explicatives et des effets fixes individuels dans un cadre non-linéaire -qui font qu'elle n'apporte pas elle non plus une confirmation robuste de l'hypothèse de persistance.

Ces difficultés seraient moindres avec des données de meilleure qualité. D'une part il serait souhaitable de disposer d'un panel comprenant plus d'individus. En effet, on pourrait alors construire des groupes occupationnels plus fins et ainsi mieux mesurer la mobilité occupationnelle et limiter les effets de frontières mentionnés au début de cet article. Il serait également possible de mener une analyse totalement différenciée par sexe qui se justifie par les comportements très distincts des hommes et des femmes sur le marché du travail péruvien. D'autre part, une base de données panel incluant plus de points dans le temps permettrait d'avoir recours aux techniques de l'économétrie de panel pour traiter les problèmes d'endogénéité et d'effets fixes. En outre, il aurait été préférable de disposer d'un échantillon national pour prendre en compte la mobilité occupationnelle générée par les flux migratoires. Enfin, il aurait été intéressant de tester cette hypothèse sur une période qui ne soit pas caractérisée par une importante crise économique. En effet, plusieurs auteurs ont montré qu'en période de crise, l'influence de la famille prend une plus grande importance qu'en période d'expansion. Au Pérou, Benavides (2002) met en avant une plus faible mobilité intergénérationnelle en période de crise économique. De même, Coury (2000) constate qu'à Madagascar, en période de crise, le capital social se substitue au capital scolaire dans l'accès aux positions salariales supérieures lors d'un premier

emploi. Elle explique ce résultat par l’informalisation des procédures d’embauches dans l’ensemble des segments du marché du travail comme conséquence de la crise économique.

En conclusion, les données à notre disposition, pourtant les seules au Pérou permettant de mener ce type d’analyse, ainsi que les difficultés techniques que soulève la problématique étudiée ne nous ont pas permis d’affirmer catégoriquement que l’origine sociale a un effet persistant sur la mobilité occupationnelle. Cependant, cette étude a le mérite de relancer le débat sur ce thème important dans un pays aussi inégalitaire que le Pérou, débat qui, à quelques exceptions près, s’était éteint depuis les années soixante-dix.

## REFERENCES BIBLIOGRAPHIQUES

- Agresti A. (1990), *Categorical Data Analysis*, New York, Wiley.
- Akerlof G. (1997), « Social Distance and Social Decisions », *Econometrica*, Vol.65, n°5, pp.1005-1027
- Altamirano T. (1988), « Cultura andina y pobreza urbana : Aymaras en Lima », Lima, Pontifica Universidad Católica del Perú, Fondo Editorial.
- Auriat N. (1996), « Les défaillances de la mémoire humaine : aspects cognitifs des enquêtes rétrospectives », Travaux et Documents de l'INED, Cahier n. 136, Paris, PUF/INED.
- Banque Interaméricaine de Développement (2004), « Good Jobs Wanted: Labor Markets in Latin America », BID, Economic and Social Progress in Latin America, Report 2004.
- Becker G. S. et Tomes N. (1986), « Human Capital and the Rise and Fall of Families », *Journal of Labor Economics* Vol. 4 (Issue 3), p. S1-S39.
- Benavides M. (2002), « Class Mobility and Equality of Opportunities in the Context of Erratic Modernization: The Peruvian case », PhD. Dissertation. Pennsylvania State University.
- Bishop Y., Fienberg S. et Holland P. (1975), *Discrete Multivariate Analysis: Theory and Practice*, Cambridge, MIT Press.
- Blau P. et Duncan O. D. (1967), *The American Occupational Structure*, New York, Wiley.
- Bocquier P. (1996), *L'insertion et la mobilité professionnelle à Dakar*, Thèse de Doctorat de 3<sup>ème</sup> cycle, Paris, Université Paris V- René Descartes-Sorbonne.
- Boudon R. (1973), *L'inégalité des chances*, Paris, Armand Colin.
- Chacaltana J. (1993), « La movilidad ocupacional en Lima Metropolitana: una aproximación micro económica », Taller de Empleo ADEC-ATEC.
- Cohen B. et House W.J. (1996), « Labor Market Choices, Earnings and Informal Networks in Karthoum, Sudan », *Economic Development and Cultural Change*, Vol 25, pp : 289-294.
- Coury D. (2000), *De l'Ecole au premier emploi : le poids de l'éducation et de l'origine sociale au cours du temps à Antananarivo*, Thèse de Doctorat de 3<sup>ème</sup> cycle. Paris, Institut d'Etudes Politiques, p. 488.
- Cuanto (1992), *Perú in números : 1992*. Lima, Cuanto S.A.
- Deming W. et Stephan F. (1940), « On a Least Squares Adjustment of a Sampled Frequency Table when the Expected Marginal Totals are Known », *Annals of mathematical statistics*, 11, 427-444.
- Galor O. et Sicherman N. (1990), « A theory of career mobility », *Journal of Political Economy*, Vol 98, p.169-192.
- Glewwe P. et Hall G. (1994), « Poverty, Inequality, and Living Standards during Unorthodox Adjustment: The Case of Peru, 1985-1990 », *Economic Development and Cultural Change*, Vol 42 (4), p.689-717.
- Granovetter M. (1974), *Getting a Job: a Study of Contacts and Careers*, Cambridge, Harvard University Press.
- Gravot P. (1995), *Economie de l'éducation*. Paris, Ed. Economica.
- Goldthorpe J.H. (1980), *Social Mobility and Class Structure in Modern Britain*, Oxford, Clarendon Press.
- Gouriéroux C. (1984), *Econométrie des variables qualitatives*, Paris, Ed. Economica.
- Goux D. et Maurin E. (1997), « Meritocracy and Social Heredity in France: Some Aspects and Trends », *European Sociological Review*, Vol.13 (2) pp.159-177.

- Goux D. et Maurin E. (1998), « From Education to First Job: the French Case » in Y. Shavit et W. Müller (eds.) *From School to Work*, Oxford, Oxford University Press, pp. 103-141.
- Grimm M. et Bonneuil N. (2001), « Labour Market Participation of French Women over the Life Cycle, 1935 to 1990 », *European Journal of Population*, Vol. 17(3).
- Haberman S. (1978), *Analysis of Quantitative Data. Volume 1 Introductory Topics*, San Diego, Academic Press.
- Haveman R. et Wolfe B. (1995), « The Determinants of Children's Attainments: a Review of Methods and Finding », *Journal of Economic Literature* Vol. 33, p. 1829-1878.
- Herrera J. (1999), "Ajuste económico, desigualdad y movilidad", in R.Webb et M.Ventocilla (eds.), *Pobreza y economía social. Análisis de una encuesta (ENNIV-1997)*, USAID, CUANTO et UNICEF, pp 101-142.
- Hout M. et Rosen S. (1999), « Self-Employment, Family Background, and Race », NBER, Working Paper, n°7344.
- Jacoby H. G. (1994), « Borrowing Constraints and Progress Through School: Evidence from Peru », *The Review of Economics and Statistics*, Vol. 76 (Issue 1) : p. 151-160.
- Kouamé A. et Gueye A. (2000), « Genre, éducation et accès au premier emploi : le cas de la ville d'Abidjan », CEPED, Les Dossiers du CEPED, n°56, 52p, Paris.
- Lam D. et Schoeni R. (1993), « Effects of Family Background on Earnings and Returns to Schooling : Evidence from Brazil », *The Journal of Political Economy*, Vol. 101(4), p.710-740.
- Maddala G.S. (1983), *Limited-dependent and Quantitative Variables in Econometrics*, Cambridge, Cambridge University Press.
- Merton R. (1953), *Reference Group Theory and Social Mobility. Class, Status and Power*. S. M. L. R. Bendix, New York, Free Press.
- Müller W. et Shavit Y. (1998), *From School to Work : a Comparative Study of Educational Qualifications and Occupational Destinations*, Oxford, Oxford University Press.
- Piketty T. (1998), « Self-fulfilling Beliefs about Social Status », *Journal of Public Economics*, Vol. 70, p. 115-132.
- Rodriguez Cuba J. (1995), « De profesional a taxista: el mercado laboral de técnicos y profesionales en los 90 », ADEC-ATC.
- Stiglitz J. (2000), « Reflections on Mobility and Social Justice, Economic Efficiency, and Individual Responsibility » in N. G. Birdsall *New Markets, new opportunities?: Economic And Social Mobility in changing World*, C. Washington D.C.
- Saavedra J. (1997), « Quiénes ganan y quiénes pierden con una reforma estructural: cambios en la dispersión de ingresos según educación, experiencia y género en el Perú urbano », GRADE, Notas para el Debate, n°14.
- Sicherman N. (1990), « Education and Occupational Mobility », *Economics of Education Review*, Vol. 9, 163-179.
- Suarez F. (1975), « La movilidad ocupacional en Lima Metropolitana », Ministerio de Trabajo, Lima, Perou.
- Thélot C. (1990), « L'erreur de mémoire sur la situation professionnelle », INSEE, Les Collections de l'INSEE : pp. 175-189.
- Zimmerman D. (1992), « Regression Toward Mediocrity in Economic Status », *American Economic Review*, Vol. 82, p. 409-29.

## ANNEXES

### Annexe 1 : Les limites du panel

#### *Evaluation de la qualité du panel : un effet de sélection ?*

Il y a des différences significatives entre les individus du panel et les autres individus vivant à Lima en 1985 (tableau A). Dans l'échantillon panel, l'âge moyen est relativement plus élevé et la population est plus féminine que dans l'autre échantillon. De plus, on y trouve relativement plus d'individus chefs de ménage ou conjoints et moins d'autres parents. Les ménages sont de plus grande taille. En outre, les individus du panel ont un niveau d'éducation moyen inférieur – une proportion plus importante à un niveau primaire, une proportion plus faible un niveau supérieur et leur nombre d'années d'éducation est relativement plus faible- que les autres Liméniens. Quant à leur situation sur le marché du travail, il est important de prendre en compte que la proportion d'actifs appartenant au groupe 1 est plus faible dans l'échantillon panel, celle d'actifs appartenant au groupe 3 plus élevée. De même, les individus du panel sont plus souvent des indépendants et moins souvent des salariés, ils sont plus souvent informels et exercent plus fréquemment une activité secondaire. En conclusion, les individus du panel ont des conditions de travail moins favorables, ils ont une dotation plus faible en capital humain et semblent avoir des niveaux de vie moins élevés -accès plus réduit au crédit et à l'épargne-. Enfin, pour ce qui est des caractéristiques liées à l'origine sociale, il semblerait que les individus du panel aient une origine sociale relativement moins élevée. En effet, ces individus ont des parents ayant des niveaux d'éducation plus faibles. Ils ont une probabilité plus élevée d'être de migrants en provenance des villes de la Sierra et de la Selva. Enfin, une proportion plus faible d'entre eux a achevé sa scolarité dans une école privée.

#### *Les limites d'un panel urbain ?*

Un reproche qui peut être fait à la base de donnée utilisée est qu'elle ne couvre que Lima. L'étude de la mobilité professionnelle urbaine peut alors être biaisée car elle ne prend pas en compte les flux migratoires vers ou hors Lima.

On peut penser par exemple qu'un individu qui connaît une mobilité descendante à Lima migrera pour tenter sa chance ailleurs. Ainsi, on sous-estimerait la mobilité descendante au profit de la mobilité ascendante. L'enquête ENNIV 1985/86 qui a une couverture nationale, permet d'étudier ces flux migratoires et d'évaluer ce biais mais uniquement pour les personnes nées à Lima. On observe que seulement 6 % de ces dernières ont quitté Lima Métropolitaine définitivement. De plus, parmi ces personnes, 80 % avaient moins de 20 ans lorsqu'ils ont migré hors de Lima et 83 % ont migré pour des raisons familiales. Cependant, si on a pu vérifier que peu de liméniens de naissance ont quitté la ville par suite de mobilité descendante, on ne sait rien sur les migrants venus à Lima tenter leurs chances et qui ne parvenant pas à s'insérer retournent dans leur région d'origine<sup>46</sup>.

Cependant, il existe une autre source de biais possible qui pourrait compenser le biais précédemment cité puisque son effet est inverse : il tendrait à sous-estimer la mobilité ascendante. Ce biais provient de la non-prise en compte des arrivées à Lima de migrants entre 1985 et 1990. Les migrants représentent 51 % de la population liménienne en 1985, 15 % d'entre eux ont migré en étant âgé de plus de 20 ans et 20 % d'entre eux vivent à Lima depuis moins de 5 ans, soit 10 % de la population de Lima. De plus, entre 1985 et 1990, les campagnes andines ont énormément souffert du terrorisme et de sa répression par les forces armées ce qui a conduit bon nombre de familles, généralement d'agriculteurs, à migrer vers Lima. Arrivées en ville, ces familles se sont reconverties vers des activités non-agricoles, connaissant dans certains cas une certaine mobilité ascendante.

---

<sup>46</sup> L'étude anthropologique d'Altamirano (1988) sur la population migrante aymara apporte des éclairages sur les migrations de retour et leurs raisons. Parmi la population étudiée (les migrants de retour à Vilquechico), la raison principale des retours est l'échec dans la ville. Ces personnes ont migré dans l'espoir d'une mobilité ascendante par le travail ou par l'école (pour eux même ou pour leurs enfants) et d'un meilleur accès à la médecine et ont été déçus dans leur attentes pour plusieurs raisons : un accès à l'école et à la santé qui s'est fortement dégradé en partie du fait de la surpopulation urbaine, un niveau élevé de sous-emploi et une forte instabilité économique. Ce retour est facilité par un fort attachement à la communauté d'origine puisque 90 % des migrants interrogés à Lima par T. Altamirano sont prêts à retourner dans leur région si les conditions de vie s'y améliorent. Selon l'auteur, le processus de retour se serait intensifié depuis 1976, début de la dégradation de la situation économique.

**Tableau A : Analyse descriptive des caractéristiques des individus, échantillons panel et non-panel en 1985.**

Variables	Echantillons	
	No Panel	Panel
<b>Caractéristiques individuelles générales</b>		
<b>Age</b>	39,1	41,4 ***
<b>Sexe (%)</b>		
- hommes	49,5	45,9 ***
- femmes	50,5	54,1 ***
<b>Statut dans le ménage</b>		
Chef de ménage	38,6	44,9 ***
Conjoint	27,4	37,2 ***
Enfant du ménage	16,7	12,8
Autres parents	17,3	5,1 ***
<b>Taille du ménage</b>	5,5	5,8 *
<b>Statut civil (%)</b>		
- vivant seul	31,7	25,1
- vivant en couple	68,3	74,9
<b>Niveau d'éducation (%)</b>		
- sans éducation	4,1	4,7
- éducation primaire	26,0	37,1 ***
- éducation secondaire	45,0	40,7 *
- universitaire et autres études du supérieur	24,9	17,6 ***
<b>Nombre d'années d'éducation</b>	9,8	8,5 ***
<b>Toujours étudiant (%)</b>	8,0	7,2
<b>Caractéristiques individuelles sur le marché du travail</b>		
<b>Groupes d'occupation</b>		
- groupe 1	19,9	15,1 **
- groupe 2	48,2	47,5
- groupe 3	31,9	37,5 **
<b>Situation sur le marché du travail (%)</b>		
- employé	76,3	78,6
- chômeur	5,3	4,5
- inactif	18,5	17,0
<b>Secteur d'emploi (%)</b>		
- primaire	7,6	9,5
- secondaire	22,6	20,3
- tertiaire	69,8	70,2
<b>Division institutionnelle (%)</b>		
- indépendants	40,4	48,7 ***
- salariés du secteur public	19,8	18,7
- salariés du secteur privé	39,7	32,6 ***
<b>Nombre d'heures travaillées pendant la semaine</b>	33,1	36,8 ***
<b>Dispose d'un contrat de travail (%)</b>	29,4	25,2
<b>Syndicat dans l'entreprise (%)</b>	39,2	42,9
<b>Couvert par la sécurité sociale (%)</b>	66,2	67,0
<b>Travaille dans le secteur informel (%)</b>	43,3	51,7 ***
<b>Dispose d'un emploi secondaire (%)</b>	5,7	9,5 ***
<b>Dispose d'un accès au crédit (%)</b>	26,6	20,4 **
<b>Dispose d'un accès à l'épargne (%)</b>	51,9	44,4 **
<b>Caractéristiques liées à l'origine sociale</b>		
<b>Niveau d'éducation du père (%)</b>		
- sans éducation	11,4	14,5 *
- éducation primaire	52,0	58,0 *
- éducation secondaire ou supérieure	36,6	27,5 ***
<b>Activité du père (%)</b>		
- Professionnel ou employé dans l'administration	18,0	15,1
- autre occupation hors l'agriculture	46,9	46,4
- agriculture	35,1	38,4
<b>Niveau d'éducation de la mère (%)</b>		
- sans éducation	27,3	37,8 ***
- éducation primaire	53,1	49,1
- éducation secondaire ou supérieure	19,5	13,1 **
<b>Activité de la mère (%)</b>		
- Professionnel ou employé dans l'administration	7,4	5,9
- autre occupation hors l'agriculture	50,4	47,2
- agriculture	42,2	46,9
<b>Région de naissance (%)</b>		
- Sierra	37,1	42,4 **
- Costa	59,1	52,6 ***
- Selva	2,9	4,4 *
<b>Lieu de naissance (%)</b>		
- La ville	42,2	50,1 ***
- Autres	57,9	49,9 ***
<b>Type d'école (%)</b>		
- Publique	78,7	86,9 ***
- Privée	21,3	13,1 ***
<b>Taille de l'échantillon</b>	<b>1574</b>	<b>1207</b>

Source : Panel 1985-1990 et ENNIV de 1985, construction des auteurs.

Notes : \* Tests de différences entre les individus du panel et le reste. \* différence significative au seuil de 10 %, \*\* de 5 % et \*\*\* de 1 %.

## Annexe 2 : La construction de la classification des catégories occupationnelles

L'occupation des enquêtés est codifiée, dans les deux enquêtes utilisées à partir de la classification internationale type des professions élaborée par l'OIT, comprenant 99 catégories.

### La classification hiérarchique ascendante

Le principe de la classification hiérarchique ascendante - *cluster analysis* - est de créer des groupes tels que chaque groupe contienne les occupations les plus semblables et que les groupes soient les plus éloignés les uns des autres. La méthode d'association retenue est agglomérative et hiérarchique : on considère chaque occupation comme un groupe (99 groupes de taille 1) puis les deux occupations les plus proches- ou similaires- sont agrégées (98 groupes, un de taille 2 et les autres de taille 1) et ainsi de suite jusqu'à ce que toutes les catégories appartiennent à un même groupe. La mesure retenue de similarité  $L2$  entre deux catégories  $i$  et  $j$  est la distance euclidienne entre ces deux catégories :

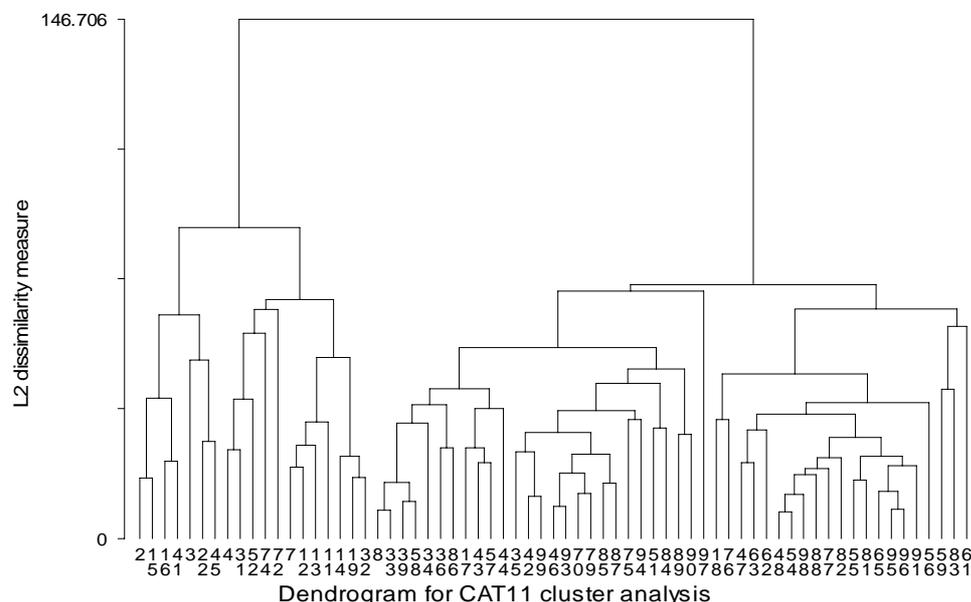
$$L2 = \sqrt{\sum_{k=1}^p (x_{ki} - x_{kj})^2} \text{ où } k \text{ est une des } p \text{ variables utilisées pour effectuer la classification. Ces}$$

variables sont ici le niveau moyen et le niveau médian de la consommation mensuelle par tête, le pourcentage de ménages épargnant au sein de chaque occupation et le nombre moyen d'années d'éducation des personnes exerçant cette occupation<sup>47</sup>. Chaque variable est considérée avec la même pondération<sup>48</sup>.

Dans la comparaison d'un groupe à un autre est prise en compte la mesure de similarité moyenne au sein de chaque groupe (*cluster averagelinkage*).

Le résultat de la classification hiérarchique ascendante est représenté par le dendrogramme suivant :

### Graphique A : Dendrogramme de la classification hiérarchique ascendante



**Lecture :** En abscisse sont représentées les occupations (ex : de gauche à droite, occupation 2, 15, 16, 41, 3, 22...), en ordonnée l'indice de similarité.

<sup>47</sup> Une des faiblesses de cette classification est que les variables qui servent à hiérarchiser les groupes sont calculées à partir de données d'enquête -LSMS 1985/86- dont la représentativité n'est pas forcément assurée au niveau des occupations non agrégées, i.e. des 99 occupations. Pour bien faire, il aurait fallu calculer ces variables à partir d'un recensement. Cependant, dans les recensements, nous n'avons pas d'informations sur le revenu que procurent les occupations.

<sup>48</sup> Dans un approfondissement futur de ce travail, il serait intéressant d'analyser la sensibilité de nos résultats au choix méthodologiques faits dans la construction des groupes d'occupations (choix des critères, des pondérations attribuées à chaque critère, à la mesure de similarité retenue dans l'analyse par cluster,...)

Dans ce dendrogramme, deux principaux groupes occupationnels se distinguent. Le premier groupe, le plus hétérogène, va de la catégorie occupationnelle 2 à 32 (abscisse du dendrogramme). Le second groupe agrège les autres catégories, *ie* les catégories 8 à 61 sur l'abscisse du dendrogramme. Pour une analyse plus fine, on souhaite désagréger ces deux groupes afin d'obtenir quatre groupes. Cependant, la désagrégation du premier groupe conduirait à créer deux sous-groupes ayant un effectif trop faible pour mener une analyse économétrique<sup>49</sup>. C'est pourquoi seul le second groupe est désagrégé. On obtient alors trois groupes, le groupe 1 comprenant les catégories 2 à 32, le groupe 2 les catégories 8 à 97 et le groupe 3 les catégories les plus défavorisées à savoir les catégories 18 à 61. Par construction, le groupe 1 est donc beaucoup plus hétérogène que ne le sont les groupes 2 et 3. De plus, les groupes 2 et 3 sont plus proches qu'ils ne le sont chacun du groupe 1 (d'après l'indice de similarité).

### *Les groupes occupationnels retenus*

L'analyse par cluster nous a conduit à retenir trois groupes. Les caractéristiques de ces groupes sont présentées dans le tableau B

**Tableau B : Caractéristiques des trois groupes occupationnels retenus**

Exemple de catégories incluses dans chaque groupe	Consommation mensuelle du ménage par tête (intis)			Part des ménages épargnant (en %)	Nombre d'années d'éducation moyen (%)	Nombre d'obs.		
	Conso moyenne	Conso médiane	Ecart type/Moy.					
<b>Groupe 1</b>	Architecte/ingénieur Journaliste Membre du gouvernement législatif Fonctionnaire supérieur Médecin/dentiste Fonctionnaire moyenne catégorie Agent administration publique Comptable Gérant et administrateur Technicien métallurgique/ chimique	Scientifique Manager/directeur Economiste Avocat Professeur	1064,5	799,2	1,05	66,7	14,3	375
<b>Groupe 2</b>	Assistant médecin Secrétaire Conducteur Personnel administratif Agent technique Personnel de sécurité Tailleur	Artiste Employé de comptabilité Vendeur Commerçant Coiffeur Contre maître Ouvrier spécialisé	603	477,2	0,93	50	9,7	981
<b>Groupe 3</b>	Vendeur ambulant Blanchisseur Pêcheur	Cuisinier Exploitant agricole Ouvrier non spécialisé	463,4	341,2	1,03	37,8	6,6	715

Source : Enquête LSMS 1985/86.

Champ : les 26-64 ans vivant à Lima

### *Discussion sur l'hypothèse de la permanence des groupes occupationnels entre 1985 et 1990.*

Même si l'intervalle de temps entre 1985 et 1990 est réduit, les chocs économiques importants qu'a connu le Pérou pendant cette période font qu'il est légitime de se demander si réellement appartenir à l'un des trois groupes procure le même statut social en 1985 qu'en 1990, relativement aux autres groupes.

Herrera (1999) s'intéresse à la mobilité économique sur la période 1985-1990 en utilisant un panel construit à partir des mêmes enquêtes que celles de cette analyse. Il montre que la mobilité

<sup>49</sup> En effet, le premier sous-groupe réunissant les catégories 2 à 45 sur l'abscisse du dendrogramme ne comprendrait que 120 observations.

économique des ménages, définie comme la mobilité entre les quintiles de consommation, a été de grande ampleur sur cette période<sup>50</sup>.

Cette forte mobilité économique, qui n'est pas forcément concomitante à une mobilité occupationnelle, laisse penser que la hiérarchie des groupes occupationnels en terme de niveau de vie s'est peut être modifiée<sup>51</sup>. Si c'est le cas, alors l'analyse de la mobilité en terme hiérarchique (ascendante et descendante) serait biaisée. On ne peut cependant le tester directement à partir de nos données, du fait de la mauvaise qualité de la mesure du revenu dans l'enquête de 1990 (cf. note 20).

Plusieurs constatations font penser que la distance entre les groupes s'est réduite entre les deux dates. Premièrement, Glewwe et Hall (1994) montrent que sur la période, les ménages dont le chef travaille en tant qu'indépendant connaissent une diminution plus faible de leur consommation que celle des employés du secteur privé ou du secteur public -51 % *versus* 56 % et 56 %-. Or le groupe 3 comprend la plus forte part d'indépendants (66 % *versus* 26 % pour le groupe 1, 42 % pour le groupe 2). La distance entre le groupe 3 et les deux autres groupes serait donc plus faible en 1990 qu'en 1985. Deuxièmement, par rapport à 1986, les salaires réels dans le secteur privé ont perdu la moitié de leur valeur en 1990, ceux du secteur public 80 % (CUANTO, 1992). Or la part des travailleurs dans le secteur public est beaucoup plus importante dans le groupe 1 que dans le groupe 2 ou 3 (45 % *versus* 18 % et 9 % pour le groupe 2 et 3). Le groupe 1 et 2 se seraient donc rapprochés. Cependant, une troisième constatation va dans le sens contraire aux deux autres : Glewwe et Hall (1994) montrent qu'entre 1985 et 1990, les ménages dont le chef à un niveau d'éducation ne dépassant pas le primaire voient leur consommation diminuer de manière plus drastique que les ménages dont le chef à un niveau d'éducation supérieur. Etant donné que la part des individus dont le niveau d'éducation ne dépasse pas le primaire est de 63 % dans le groupe 3, de seulement 4% dans le groupe 1 et de 33 % dans le groupe 2<sup>52</sup>, on pourrait penser que le fossé entre les trois groupes s'est creusé sur la période.

Si l'écart entre les groupes s'est probablement réduit entre 1985 et 1990, il semble que la hiérarchie entre les groupes soit restée la même. En effet, selon différents auteurs, la hiérarchie en terme de revenu des occupations est restée la même, que les occupations soient appréhendées par leur niveau d'éducation (Saavedra, 1997), leur statut<sup>53</sup> (Glewwe et Hall, 1994) ou qu'elles soient agrégées selon un critère différent du notre<sup>54</sup> (Rodríguez, 1995).

---

<sup>50</sup> Le taux d'immobilité est de 37 %, la part de ceux se déplaçant d'un quintile est 40 % et le saut moyen entre 1985 et 1990 représente 61 % du saut maximum que l'on obtient en cas de mobilité parfaite.

<sup>51</sup> Elle ne s'est pas modifiée en terme de niveau d'éducation : les niveaux d'éducation moyens à l'intérieur de chaque groupe ne sont pas significativement différents entre 1985 et 1990.

<sup>52</sup> La part des individus ayant un niveau d'éducation ne dépassant pas le primaire à l'intérieur de chaque groupe en 1985 n'est pas significativement différente de celle en 1990.

<sup>53</sup> Occupation indépendante vs. salariée.

<sup>54</sup> Les occupations sont agrégées en 4 groupes (professions qualifiées –« *profesionales* »-, employés, ouvriers qualifiés, vendeurs ambulants).

### Annexe 3 : Présentation des modèles log-linéaires<sup>55</sup>

L'objectif de la modélisation log-linéaire est d'éprouver des hypothèses sur les interactions –ou associations statistiques- entre des variables. Pour ce faire, on cherche à reconstituer la table de contingence formée par ces variables à travers plusieurs modèles emboîtés. Chaque modèle fait une hypothèse différente sur les interactions entre les variables. Il estime la fréquence observée de chacune des cellules de la table de contingence en supposant plus ou moins d'interactions entre les variables. L'approche par les modèles log-linéaires est hiérarchique. Un premier modèle, appelé modèle de référence, ne comporte aucune interaction. Ce modèle de référence est emboîté dans un second qui comprend une interaction supplémentaire et ainsi de suite jusqu'au modèle saturé qui inclut toutes les interactions possibles entre les variables –qui n'a donc plus de degré de liberté- et dans lequel sont contenus tous les modèles précédents. Les interactions entre les variables retenues seront celles du modèle qui reproduit le mieux les données observées et avec la plus grande parcimonie de paramètres. Différents indicateurs rendent compte des qualités prédictives et de la parcimonie de chacun des modèles. Ils sont présentés en annexe 4.

La différence principale entre un modèle logit et un modèle log-linéaire est que le modèle log-linéaire considère toutes les variables comme des variables de réponses, c'est-à-dire qu'aucune distinction n'est faite entre variable dépendante et variables indépendantes. Autrement dit, les modèles log-linéaires ne font que décrire les associations statistiques entre les variables, sans supposer une quelconque relation de causalité. Il est cependant possible de considérer un modèle logit polytomique univarié comme un modèle log-linéaire, à condition d'imposer des contraintes linéaires sur les paramètres du modèle logit<sup>56</sup>.

Pour illustrer la méthodologie des modèles log-linéaires, nous nous proposons ici de détailler l'approche pour une table de contingence à deux dimensions. Tester les interactions entre les variables pour une table à deux dimensions revient à tester l'indépendance entre les deux variables qui forment cette table. Supposons qu'en ligne soit représentée l'activité du père O, à  $I$  modalités, et en colonne, l'activité de son enfant E à  $J$  modalités.

Si les deux variables sont indépendantes, alors, par définition :

$$n_{ij} = \frac{\sum_j n_{ij} \sum_i n_{ij}}{n} = \frac{n_{i.} n_{.j}}{n} \text{ où } n_{ij} \text{ est la fréquence de la cellule formée par la ligne } i \text{ et la colonne } j, n$$

l'effectif total de la table. L'expression, sous forme logarithmique équivaut à :

$$\ln n_{ij} = -\ln n + \ln n_{i.} + \ln n_{.j}$$

On cherche maintenant à tester grâce aux modèles log-linéaires quelles sont les interactions entre les deux variables. Le modèle log-linéaire de référence est celui qui n'inclut aucune interaction entre les variables. Il s'écrit :

$$\ln(F_{ij}) = \mu + \lambda_i^O + \lambda_j^E.$$

où  $F_{ij}$  est la fréquence estimée de la cellule  $(i, j)$ . Elle est égale, sous l'hypothèse d'indépendance, à :

$$F_{ij} = \frac{n_{i.} n_{.j}}{n}.$$

<sup>55</sup> Pour plus de détails, se référer à Bishop et al. (1975), Agresti (1990).

<sup>56</sup> Pour une présentation du lien formel entre un modèle log-linéaire et un modèle logit polytomique univarié, se référer à Gouriéroux (1984), p. 148-149.

Le paramètre  $\mu$  est une constante. Il est la moyenne des logarithmes de l'ensemble des fréquences estimées de la table :  $\mu = \frac{1}{IJ} \sum_i^I \sum_j^J \ln(F_{ij})$

Le paramètre  $\lambda_i^O$  est l'écart à la moyenne  $\mu$  dont est responsable la variable  $O$  :

$$\lambda_i^O = \frac{1}{J} \sum_j^J \ln(F_{ij}) - \mu,$$

le paramètre  $\lambda_j^E$  l'écart à la moyenne  $\mu$  dont est responsable la variable  $E$  :

$$\lambda_j^E = \frac{1}{I} \sum_i^I \ln(F_{ij}) - \mu.$$

Si les deux variables sont indépendantes, alors le modèle de référence reproduit parfaitement la table observée. On obtient :

$$F_{ij} = n_{ij} \quad \forall i, j ; \mu = \ln n ; \lambda_i^O = \ln n_{i.} - \ln n ; \lambda_j^E = \ln n_{.j} - \ln n$$

Si ce n'est pas le cas, un test par la statistique de vraisemblance conduit à rejeter l'hypothèse selon laquelle les fréquences estimées  $F_{ij}$  ne sont pas significativement différentes des fréquences observées  $n_{ij}$  (cf. annexe 4). Il faut alors introduire un nouveau paramètre dans le modèle, une interaction entre  $O$  et  $E$ ,  $\lambda_{ij}^{OE}$  qui est le logarithme du rapport entre la vraie fréquence de la cellule et la fréquence qu'elle aurait si les deux variables étaient indépendantes :  $\lambda_{ij}^{OE} = \ln(F_{ij}) - \mu - \lambda_i^O + \lambda_j^E$ . Le modèle devient alors le modèle saturé :

$$\ln(F_{ij}) = \mu + \lambda_i^O + \lambda_j^E + \lambda_{ij}^{OE}.$$

Pour une table à trois dimensions, on cherche à tester les différentes interactions entre les variables. Soit une table de contingence formée par trois variables  $O$  à  $I$  modalités,  $E$  à  $J$  modalités, et  $D$  à  $K$  modalités.

Le modèle de référence qui suppose l'indépendance entre toutes les variables est le suivant :

$$\ln(F_{ijk}) = \mu + \lambda_i^O + \lambda_j^E + \lambda_k^D$$

Dans ce modèle, la fréquence estimée est :  $F_{ijk} = \frac{n_{i..} n_{.j.} n_{..k}}{n^2}$  et les paramètres sont obtenus à partir des

formules suivantes :  $\mu = \frac{1}{IJK} \sum_i^I \sum_j^J \sum_k^K \ln(F_{ijk})$ ,  $\lambda_i^O = \frac{1}{JK} \sum_j^J \sum_k^K \ln(F_{ijk}) - \mu, \dots$

Si ce modèle ne reproduit pas correctement les données observées, on estime le modèle de référence auquel on ajoute une interaction partielle entre deux variables, étant donnée la troisième. Par exemple, on suppose qu'il y a une association statistique entre l'origine et l'éducation, étant donnée la destinée,  $\lambda_{ij}^{OE}$ . Le modèle s'écrit alors :

$$\ln(F_{ijk}) = \mu + \lambda_i^O + \lambda_j^E + \lambda_k^D + \lambda_{ij}^{OE}$$

La fréquence estimée est égale dans ce modèle à :  $F_{ij} = \frac{n_{ij} \cdot n_{..k}}{n}$  et le paramètre correspondant à cette interaction partielle est égal à :

$$\lambda_{ij}^{OE} = \frac{1}{K} \sum_k \ln(F_{ijk}) - \frac{1}{IK} \sum_i \sum_k \ln(F_{ijk}) - \frac{1}{JK} \sum_j \sum_k \ln(F_{ijk}) + \mu$$

les autres paramètres étant inchangés par rapport au modèle précédent.

Si la qualité d'ajustement du modèle incluant une seule interaction partielle est toujours insuffisante, il faut ajouter au modèle précédent une interaction partielle supplémentaire, par exemple l'interaction partielle entre l'éducation et la destinée, étant donnée l'origine,  $\lambda_{jk}^{ED}$ . Le modèle s'écrit alors :

$$\ln(F_{ij}) = \mu + \lambda_i^O + \lambda_j^E + \lambda_k^D + \lambda_{ij}^{OE} + \lambda_{jk}^{ED}$$

La fréquence estimée est égale dans ce modèle à :  $F_{ij} = \frac{n_{ij} \cdot n_{.jk}}{n_{.j}}$  et le paramètre correspondant à cette interaction partielle est, comme précédemment, égal à :

$$\lambda_{jk}^{ED} = \frac{1}{I} \sum_i \ln(F_{ijk}) - \frac{1}{IJ} \sum_i \sum_j \ln(F_{ijk}) - \frac{1}{IK} \sum_i \sum_k \ln(F_{ijk}) + \mu.$$

Si ce modèle est toujours insatisfaisant, on estime le modèle qui inclut les trois interactions partielles possibles entre les variables, soit le modèle :

$$\ln(F_{ij}) = \mu + \lambda_i^O + \lambda_j^E + \lambda_k^D + \lambda_{ij}^{OE} + \lambda_{jk}^{ED} + \lambda_{ik}^{OD}$$

Dans ce cas, la fréquence estimée ne peut être obtenue à partir d'une simple formule puisque aucune paire de variable n'est indépendante conditionnellement à la troisième variable. C'est pourquoi, la fréquence doit être estimée à l'aide d'une procédure itérative basée sur l'algorithme IPF<sup>57</sup> de Deming-Stephan (1940) ou celui de Newton-Raphson<sup>58</sup>.

<sup>57</sup> « Iterative Proportional Fitting »

<sup>58</sup> pour une présentation de ces algorithmes, se référer à Haberman (1978), p. 125-133.

#### Annexe 4 : Tests de spécification des modèles log-linéaires

La construction d'un test à partir de modèles log-linéaires implique de comparer plusieurs modèles emboîtés entre eux. Le choix d'un modèle s'opère alors à partir de deux critères *a priori* contradictoires. Le premier critère est la qualité d'ajustement du modèle aux données observées. Le second est la parcimonie du modèle, c'est-à-dire sa capacité à reproduire les données avec le moins de paramètres possible. Le modèle retenu dépendra de l'importance relative que l'on attache à chacun de ces critères.

Le modèle le plus simple (modèle de référence) qui suppose l'indépendance entre toutes les variables, est le modèle le moins apte à reproduire les données mais le plus économe en paramètres (il a le degré de liberté le plus élevé). Dans le deuxième modèle, on introduit une première interaction entre les variables. La qualité d'adéquation du modèle s'améliore mais au prix de la perte de degrés de liberté. Ce modèle contient le premier. Un troisième modèle intègre une interaction supplémentaire et comprend donc les deux précédents modèles. La démarche se répète jusqu'au modèle saturé. Ce dernier reproduit parfaitement les données mais n'a aucune valeur explicative, son degré de liberté étant nul. En partant du modèle de référence, on compare successivement les modèles deux à deux en se demandant si la meilleure qualité prédictive du modèle contenant le plus d'interactions justifie son plus grand nombre de paramètres.

Pour cela, le statisticien dispose de plusieurs indicateurs. Un premier indicateur de l'adéquation du modèle est l'indice de dissimilarité,  $\Delta\%$ , qui indique le pourcentage de personnes mal classées par le modèle, i.e. la proportion de l'effectif total qui devrait être changée de cellule dans la table de contingence estimée pour que celle-ci soit égale à la table observée. Il s'écrit :

$$\Delta\% = \frac{1}{2n} \sum_i \sum_j \sum_k |n_{ijk} - F_{ijk}|$$

La faiblesse de cet indicateur est qu'il n'intègre pas le critère de parcimonie du modèle. Il n'en est pas de même pour la statistique du rapport de vraisemblance,  $G^2$ , qui permet de tester l'adéquation du modèle aux données empiriques, en prenant en compte le nombre de degrés de liberté. Elle se définit ainsi :

$$G^2 = 2 \sum_i \sum_j \sum_k n_{ijk} \ln \left( \frac{n_{ijk}}{F_{ijk}} \right)$$

$G^2$  suit asymptotiquement une loi de  $\chi^2$  au nombre de degrés de liberté égal à celui du modèle, *ddl*. Si  $p < 0,05$ , alors l'hypothèse selon laquelle les données estimées par le modèle ne sont pas significativement différentes des données empiriques est rejetée au seuil de 5%. Cependant, cette statistique est sensible à la taille de l'échantillon : plus l'échantillon est grand, plus le test construit à partir de  $G^2$  conduit à rejeter le modèle.

A partir de la statistique  $G^2$  est construit l'indicateur  $rG^2$ . Si l'on compare un modèle de référence (modèle (1)) avec un modèle comprenant plus d'interactions (modèle (2)), soit plus de paramètres, la valeur de  $rG^2$  est le pourcentage de la distance entre le modèle (1) et les données observées que le modèle (2) est capable d'expliquer, c'est-à-dire une mesure de la supériorité d'adéquation du modèle (2) aux données observées par rapport au modèle (1). Il s'écrit comme suit :

$rG^2 = 100 * \left( 1 - \frac{G^2_{(2)}}{G^2_{(1)}} \right)$  où  $G^2_{(1)}$  et  $G^2_{(2)}$  sont les statistiques du rapport de vraisemblance du modèle (1) et du modèle (2) respectivement.

Puisque les modèles sont emboîtés, il est possible de choisir le meilleur modèle en testant l'opportunité de rajouter des paramètres supplémentaires pour mieux reproduire les données observées. Autrement dit, un modèle (B) à  $d-l$  degrés de liberté ( $l > 0$ ) sera préféré à un modèle (A) à  $d$  degrés de liberté -le modèle (B) contient le modèle (A)-, si le gain de vraisemblance qu'il génère grâce à l'ajout de paramètres justifie la perte de  $l$  degrés de liberté.

Pour choisir, nous disposons de deux méthodes : des tests de  $\chi^2$  entre les deux modèles et la comparaison des indicateurs BIC (Bayesian Information Criterion) de chacun des modèles.

Le principe des tests de  $\chi^2$  est le suivant :

Puisque la différence des statistiques du rapport de vraisemblance entre le modèle (B) et le modèle (A),  $G_{(A)}^2 - G_{(B)}^2$ , suit une loi de  $\chi^2$  à  $l$  degrés de liberté, le modèle (B) sera préféré au modèle (A) si la différence  $G_{(A)}^2 - G_{(B)}^2$  est supérieure à la valeur du quantile à 95 % d'un  $\chi^2$  à  $l$  degrés de liberté.

Les tests de  $\chi^2$  accordent une importance prépondérante à la qualité d'ajustement du modèle. A l'inverse, dans la comparaison des indicateurs BIC, c'est la parcimonie du modèle qui est surtout valorisée. Ces indicateurs tiennent à la fois compte de l'adéquation du modèle, de sa parcimonie et de la taille de l'échantillon  $n$ . Ils sont définis ainsi :

$$BIC = G^2 - (ddl) \ln(n)$$

Le modèle (B) sera préféré au modèle (A) si son BIC est moindre (d'au moins 6 points). C'est pour ce modèle que l'équilibre entre adéquation du modèle et multiplicité des paramètres est le meilleur.

## Annexe 5 : Estimations des modèles 1 et 2

	Modèle 1		Modèle 2	
	Coef.	Sign.	Coef.	Sign.
<b>Equation 1: Le groupe occupationnel en 1990</b>				
<i>Variable dépendante</i>	1=Être du groupe 1 en 1990 0=Être du groupe 2 ou 3		1=Être du groupe 2 en 1990 0=Être du groupe 3	
<b>Variables sur les caractéristiques individuelles</b>				
Sexe (femmes = 1)	-0,34		-0,17	
Age	0,01		0,00	
Sexe*Age (réfèrent: avoir plus de 40 ans)				
Femme de moins de 40 ans	-0,23		0,02	
Homme de moins de 40 ans	-0,19		-0,31	
<b>Variables d'origine sociale</b>				
<i>Niveau d'éducation des parents (réfèrent: Education primaire)</i>				
Sans éducation	-0,53 *		-0,31	
Education secondaire ou supérieure	0,14		0,04	
<i>Activité du père (réfèrent: Agriculteur)</i>				
Professionnel ou employé dans l'administration	0,38		0,11	
Autres hors l'agriculture	0,07		-0,34 *	
<i>Lieu de naissance (ville=1)</i>	-0,07		0,11	
<i>Type d'école (publique = 1)</i>	-0,12		-0,23	
<b>Le groupe occupationnel en 1985</b>				
Être du groupe 1 (réfèrent: être du groupe 2 ou 3)	1,85 ***			
Être du groupe 2 (réfèrent: être du groupe 3)			1,83 ***	
<b>Constante</b>	-1,41 ***		-0,14	
<b>Equation 2: Le groupe occupationnel en 1985</b>				
<i>Variable dépendante</i>	1=Être du groupe 1 en 1985 0=Être du groupe 2 ou 3		1=Être du groupe 2 en 1985 0=Être du groupe 3	
<b>Variables sur les caractéristiques individuelles</b>				
Sexe (femmes = 1)	-0,19		0,37	
Age	-0,02		0,00	
Sexe*Age (réfèrent: avoir plus de 40 ans)				
Femme de moins de 40 ans	0,31		-0,14	
Homme de moins de 40 ans	0,25		0,35	
<b>Variables d'origine sociale</b>				
<i>Niveau d'éducation des parents (réfèrent: Education primaire)</i>				
Sans éducation	-0,15		0,02	
Education secondaire ou supérieure	-0,19		0,09	
<i>Activité du père (réfèrent: Agriculteur)</i>				
Professionnel ou employé dans l'administration	0,64		0,22	
Autres hors l'agriculture	-0,29		-0,12	
<i>Lieu de naissance (ville=1)</i>	-0,07		0,05	
<i>Type d'école (publique = 1)</i>	-0,06		-0,02	
<b>Instruments</b>				
<i>Le groupe occupationnel en 1982</i>				
Être du groupe 1 (réfèrent: être du groupe 2 ou 3)	3,91 ***		3,26 ***	
Être du groupe 2 (réfèrent: être du groupe 3)				
<i>Le groupe occupationnel du conjoint en 1982 (réfèrent: inactivité)</i>				
Être du groupe 1				
Être du groupe 2				
Être du groupe 3				
<b>Constante</b>	-1,09		-1,72 **	
<b>Rho</b>	-0,27		-0,47	
<b>Log Likelihood</b>	-257,55		-288,34	
<b>N. observations</b>	597		443	

Source: Panel LSMS 1985-1990, construction des auteurs.

Champ: Les individus de 26-64 ans en 1985 et plus exerçant une activité en 1985 et en 1990.

Note : \* coefficient significatif au seuil de 10%, \*\* au seuil de 5%, \*\*\* au seuil de 1%.