

**Direction des Études et Synthèses Économiques**

**G 2007 / 06**

**Massification et démocratisation  
de l'enseignement supérieur en France**

Valérie ALBOUY et Chloé TAVAN  
(en collaboration avec Michel DUÉE )

**Document de travail**



**Institut National de la Statistique et des Études Économiques**

# INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES

*Série des documents de travail  
de la Direction des Études et Synthèses Économiques*

**G 2007 / 06**

## **Massification et démocratisation de l'enseignement supérieur en France**

Valérie ALBOUY et Chloé TAVAN \*  
(en collaboration avec Michel DUÉE \*\*)

JUILLET 2007

Nous remercions les participants du séminaire D3E et des ateliers de la DEPP, ainsi que toutes les personnes qui nous ont aidées dans notre réflexion tout au long de ce travail, en particulier Louis-André Vallet. Nous restons responsables des erreurs qui subsisteraient.

---

\* Département des Études Économiques d'Ensemble - Division « Redistribution et Politiques Sociales »  
Timbre G210 - 15, bd Gabriel Péri - BP 100 - 92244 MALAKOFF cedex

\*\* Faisait partie du D3E lors du lancement de cette étude.

## **Massification et démocratisation de l'enseignement supérieur en France**

### **Résumé**

L'ouverture rapide du supérieur depuis le début des années 1980 et la diversification de son offre éducative, avec notamment le développement des IUT et des STS, invitent à réexaminer la question de la réduction des inégalités sociales devant l'école. La réalité de cette démocratisation est doublement contestée. D'une part, l'enseignement supérieur se serait essentiellement démocratisé au niveau de ses premiers cycles, le mouvement de démocratisation s'essouffant pour les diplômes plus élevés. D'autre part, les inégalités auraient changé de nature et concerneraient désormais la nature des études suivies. La compilation des enquêtes Emploi 1990 à 2002 permet d'étudier l'évolution fine du lien entre origine sociale et diplôme sur longue période.

L'ouverture du supérieur ne s'est pas accompagnée d'une pause dans le mouvement de démocratisation et ce, quel que soit le niveau de diplôme considéré. Nous trouvons en outre que la démocratisation a été de même ampleur pour les différents cycles de l'enseignement supérieur. Mesurée à niveau de sélection constant, afin de neutraliser l'éventuelle dévalorisation des titres scolaires, la baisse de la sélectivité sociale du supérieur est encore avérée. En revanche, l'enseignement supérieur dans son ensemble s'est moins démocratisé que le baccalauréat.

La prise en compte des différenciations « horizontales » des études, à travers les filières, conclut que, pour les filles, la phase d'expansion rapide du supérieur a été de pair avec une accentuation de la polarisation sociale de ses différentes filières : les filières les plus « bourgeoises » le sont devenues encore plus ; de même pour les filières les plus populaires. En revanche, la polarisation sociale des orientations des garçons dans le supérieur semble plutôt constante.

**Mots-clés** : inégalités sociales, enseignement supérieur, disciplines, système éducatif, catégorie socioprofessionnelle

## **Dynamics of Social Inequalities in French Higher Education**

### **Abstract**

The role of socio economic background in school attainment is a central issue. Most studies conclude that if socioeconomic background is still strongly associated with school attainment, this association has rather declined among cohorts, including for higher degrees. Opponents of the decline of educational inequalities put forward two arguments. They first argue that decline of social differences in higher degree's completion rate are mainly due to "first cycle" diplomas. They also claim that inequalities in education might be maintained through horizontal educational choices: as higher education expanded, school level achieved became less and less an issue, and the very precise nature of the diploma became more and more important.

We use data from the French Labour Force Survey collected from 1990 to 2002 to study changes in education stratification among cohorts.

Our first concern is to determine whether social inequalities have more decreased among students at lower degree levels than among those holding a higher degree. Thanks to a simulation approach, we conclude that social differences in completion rates have evenly decreased when measured at different degrees of higher education. Measuring social inequalities for constant selectivity (and not for constant diplomas) confirms that social inequalities in higher education have decreased as well. Nevertheless, our results show that the decrease of social disparities among higher education completion rates has been less pronounced than the decrease of social disparities among secondary education completion rates.

Our second concern is to determine if social polarization of different fields of studies within higher education has increased. Our results are that women's choices regarding higher education field have become more and more socially stratified, especially in the recent cohorts, when higher education has widely opened up. In the meantime, we do not observe any significant rise for men.

**Keywords**: social stratification, higher education, educational system, family background, horizontal educational choices

**Classification JEL** : I20, I21, N34

## Sommaire

Introduction .....	5
<b>I - Ouverture et transformation qualitative de l'enseignement supérieur .....</b>	<b>7</b>
<i>I.1 Générations nées entre 1930 et la fin des années 40 : une première phase de développement du supérieur</i>	8
<i>I.2 Générations 1948-1962 : la diffusion du supérieur marque le pas pour les garçons, se poursuit pour les filles</i>	8
<i>I.3 Générations 1960-1977 : l'ouverture du supérieur reprend</i>	8
<b>II - Une démocratisation à tous niveaux ? .....</b>	<b>11</b>
<i>II.1 Le débat autour de la question de la démocratisation de l'école</i>	11
<i>II.2 Comparaison de la vitesse de démocratisation des différents cycles</i>	13
II.2.1 Comment comparer la démocratisation à deux niveaux de scolarité différents ?	14
II.2.2 Les résultats	17
<i>II.3 A taux de sélection constant, le supérieur s'est démocratisé</i>	19
<b>III - Une polarisation sociale accrue des filières du supérieur ?....</b>	<b>21</b>
<i>III.1 Les inégalités horizontales, une question ancienne relativement peu étudiée</i>	21
<i>III.2 L'ouverture du supérieur s'est accompagnée d'une recomposition de ses filières</i>	21
<i>III.3 Les résultats</i>	22
<b>Bibliographie .....</b>	<b>25</b>
<b>Annexe 1 : Données et nomenclatures .....</b>	<b>27</b>
<b>Annexe 2 : Quel champ retenir en termes d'âge ? .....</b>	<b>32</b>
<b>Annexe 3 : Données de cadrage complémentaires .....</b>	<b>36</b>
<b>Annexe 4 : Méthodologie .....</b>	<b>43</b>
<b>Annexe 5 : Paramètres d'intensité des modèles d'évolution uniforme dans l'analyse par filière .....</b>	<b>48</b>
<b>Annexe 6 : Mise en œuvre pratique du test de significativité des coefficients <math>\beta_g</math> entre deux modèles différents .....</b>	<b>49</b>



## Introduction

### **L'ouverture de l'enseignement supérieur s'est accompagnée d'une transformation de son organisation**

L'enseignement supérieur en France a connu une très forte expansion quantitative à partir du début des années 1960 jusqu'à la fin des années 1990 ; depuis ses effectifs se sont stabilisés. En l'espace de quarante ans, le nombre de ses inscrits a été multiplié par sept, passant de 309 700 à plus de 2,1 millions (MEN, 2002). Ce passage, relativement rapide, d'un enseignement réservé à une élite à un enseignement de masse n'a pas été spécifique à la France. Il a concerné la plupart des pays développés, dans des proportions variables. Dans tous ces pays, cette ouverture au plus grand nombre n'a pas pu se faire sans une profonde transformation de l'organisation du système d'enseignement supérieur (Shavit, Arum, Gamoran, 2007). En France, la partie émergée de cette mutation a été la création et le développement de filières d'enseignement à finalité professionnelle. Les IUT, créés en 1966, accueillent 115 100 étudiants à la rentrée 2002. Les STS comptent à la même date 245 100 étudiants, soit presque quatre fois plus que vingt ans plus tôt. L'ouverture du supérieur a donc été de pair avec une différenciation accrue de son offre d'enseignement.

### **Une démocratisation apparente contestée**

#### *Report du seuil où s'exercent les inégalités sociales*

Comme pour l'ensemble du système éducatif, les proportions de diplômés du supérieur des différents milieux sociaux sont un peu moins dispersées dans les générations récentes (Selz, Vallet, 2006). Mais la réalité de cette démocratisation apparente est contestée par de nombreux chercheurs. Ces derniers avancent l'idée que les inégalités sociales face à l'école se seraient « simplement translatées vers le haut, quand elles ne [seraient] pas accrues » et seraient devenues « moins lisibles », pour reprendre les expressions de Piketty (Le Monde, 21/02/2006). Ainsi, pour Merle (2002), la démocratisation de l'école établie par Thélot et Vallet (2000) serait largement tributaire de la nomenclature des diplômes retenue : en n'isolant au sein du supérieur que les niveaux « baccalauréat + 1 ou 2 an(s) » et « baccalauréat + 3 ans ou plus », les auteurs ne pourraient pas rendre compte du « déplacement de ces inégalités au-delà du niveau « bac+3 » ». Le mouvement de report des inégalités observé par Duru-Bellat et Kieffer (2000) pour le secondaire concernerait désormais le supérieur.

#### *Des inégalités plus qualitatives dans la nature des études suivies*

Les inégalités auraient également changé de nature : des inégalités quantitatives dans les chances d'accès à un niveau d'enseignement donné auraient cédé la place à des inégalités plus qualitatives dans la nature des études suivies. Duru-Bellat et Van Zanten (1999) notent ainsi, pour le collège, qu'« à des inégalités dans le fait de rester ou non dans le collège pendant tout le premier cycle, se sont substituées des inégalités plus qualitatives de réussite (probabilité inégale de redoubler) et d'orientation ». De la même façon, au lycée et dans le supérieur, « le caractère plus ou moins distinctif de la filière devient davantage un enjeu social et se substitue, en partie, à la durée des études » (Merle, 1996). Les résultats de Albouy et Wanecq (2003) vont dans ce sens : alors que les troisièmes cycles se sont continûment démocratisés au cours du siècle, la base sociale de recrutement des grandes écoles s'est resserrée dans les années 1980. De même, Merle (1996) souligne qu'au sein du supérieur, la dynamique des inégalités sociales diffère selon le type d'établissement (Université, IUT, écoles d'ingénieur...). Du fait du déplacement des inégalités au fil du cursus scolaire et de leur changement de nature, Merle (2002) invite donc à

« mesurer les différences de carrière là où elles se situent, c'est-à-dire de plus en plus au niveau (...) des différentes formations de l'enseignement supérieur ».

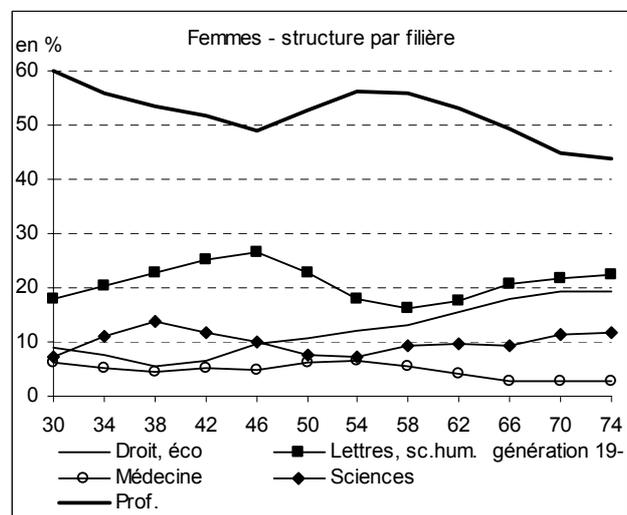
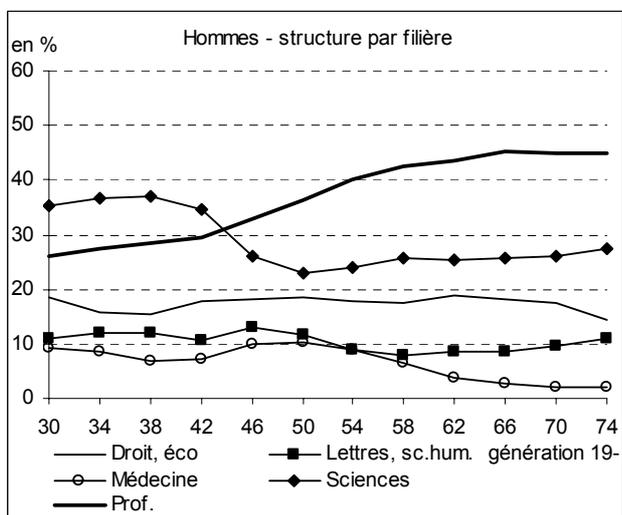
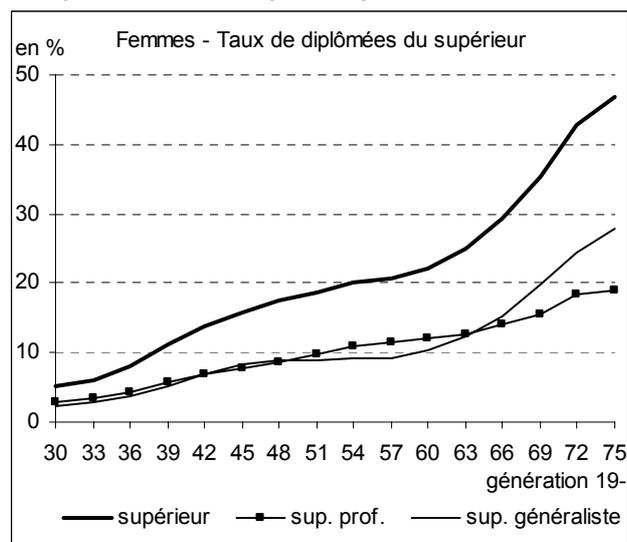
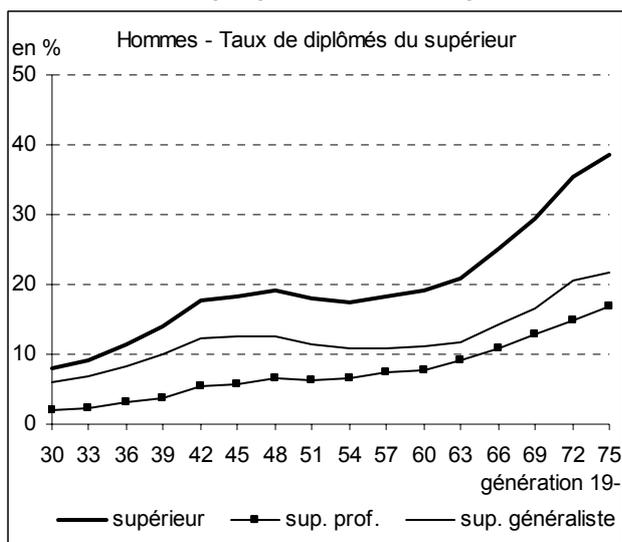
Ce texte s'inscrit dans la suite des travaux qui ont été menés en France sur le lien entre origine sociale et destinée scolaire en se proposant de regarder ce lien à un niveau plus fin, à la fois dans la dimension verticale de l'enseignement supérieur (en distinguant les trois cycles qui le composent) et dans sa dimension horizontale en isolant ses différentes filières d'enseignement (Van de Werfhorst, Luijkx, 2006). La question à laquelle nous nous proposons de répondre est la suivante : dans quelle mesure l'ouverture rapide de l'enseignement supérieur a engendré ou non un déplacement des inégalités et/ou un accroissement des inégalités sociales en son sein ? La démocratisation du supérieur a-t-elle concerné seulement le premier cycle, et notamment ses filières professionnalisantes, ou a-t-elle touché l'ensemble de ses niveaux ? Le recrutement social de ses filières s'est-il modifié ? Une exploitation des enquêtes Emploi de 1990 à 2002 permet d'apporter des éléments de réponse (Annexe 1). L'approche ne sera ici que descriptive : l'enjeu est d'enrichir la description un peu monolithique par niveau de diplôme par une approche prenant davantage en compte la diversité des filières de l'enseignement supérieur. Les limites de ces approches descriptives sont bien connues : elles ne disent rien sur les processus de formation des inégalités sociales. Elles ne permettent notamment pas de faire la part entre ce qui relève de l'auto-sélection des individus (ce qui relève purement de stratégies et de projets différents) et ce qui relève d'un choix contraint par le niveau scolaire des élèves au moment de la décision d'orientation.

Nous rappellerons dans une première partie à quel point l'enseignement supérieur s'est progressivement ouvert au cours des cinquante dernières années. Nous préciserons notamment les recompositions de la structure des enseignements qui ont accompagné cette ouverture. La seconde partie aborde la question du report de la sélection sociale à des niveaux plus élevés du cursus scolaire. Nous proposons deux manières de répondre à cette interrogation. La première est d'établir si la démocratisation a été de plus en plus faible au fur et à mesure qu'on avance dans le cursus. La seconde consiste à mesurer les disparités sociales non pas à un niveau de diplôme donné, mais à un niveau de sélection globale constant. Dans une troisième partie, nous examinerons ce que certains chercheurs nomment « les inégalités horizontales » en regardant si l'ouverture du supérieur a été de pair avec une polarisation sociale accrue de ses filières.

## I - Ouverture et transformation qualitative de l'enseignement supérieur

En l'espace de près de 50 ans, entre les générations nées au début des années 1930 et celles nées au milieu des années 1970, les études supérieures ont connu une forte expansion, à un rythme variable selon les générations (Graphiques 1). Cette diffusion est le fruit non seulement d'une diffusion des études secondaires, mais aussi d'une hausse, pour certaines générations, de la propension à continuer ses études après le baccalauréat (Annexe 3, A2).

Graphiques 1 : Taux de diplômés du supérieur et répartition de ces diplômés par filière



Source : enquêtes Emploi 1990, 1993, 1996, 1999, 2002.

Champ : personnes nées en France, âgées d'au moins 25 ans.

Lecture : 2 % des hommes nés au début des années 30 sont diplômés du supérieur professionnalisant, 6,1 % du supérieur général. Au total, 8,1 % des hommes de cette génération sont diplômés du supérieur, dont environ 35 % d'un diplôme généraliste dans les disciplines scientifiques.

### ***1.1 Générations nées entre 1930 et la fin des années 40 : une première phase de développement du supérieur***

#### **La part de diplômés du supérieur a doublé, mais reste inférieure à 20 %**

La forte croissance du nombre d'étudiants remonte aux générations nées entre le début des années 1930 et la fin des années 1940. Elle est d'abord la conséquence d'une expansion démographique liée à l'arrivée dans le système éducatif des générations du baby-boom, mais pas seulement : elle résulte aussi d'une élévation du niveau d'instruction. Les enfants de ces générations font des études plus longues que leurs aînés. Le secondaire se diffuse : 28 % de la génération 1948-50 termine ses études avec un diplôme supérieur ou égal au baccalauréat ; c'est 16 points de plus que pour ceux nés 18 ans plus tôt. Les études supérieures aussi se diffusent (Annexe 3 - A2) : ces nouveaux bacheliers poursuivent de plus en plus fréquemment des études supérieures (proportion qui est passée de 53 % à 65 % pour ces mêmes générations). Au final, la part de diplômés du supérieur triple presque sur la période ; elle passe précisément de 6,5 % à 18,4 %.

#### **Les études professionnalisantes attirent une part croissante de jeunes hommes**

Pour ces générations, les diplômés du supérieur valident pour plus d'un garçon sur deux des études scientifiques ou des formations professionnalisées courtes. Bien qu'alors essentiellement constituées de formations conduisant aux métiers très féminisés des secteurs de la santé et de l'éducation, ces études professionnalisantes attirent une proportion croissante d'étudiants masculins au détriment des études scientifiques. Pour les filles, la part des études professionnalisantes, bien qu'importante, diminue continûment au fil des générations, passant de 60 % pour la génération 1930-33 à 49 % pour la génération 1946-49. Ce sont les lettres et les sciences humaines qui bénéficient le plus de cette première phase d'expansion de l'enseignement supérieur.

### ***1.2 Générations 1948-1962 : la diffusion du supérieur marque le pas pour les garçons, se poursuit pour les filles***

Pour les générations nées entre 1948 et 1962, la hausse du niveau d'instruction marque le pas pour les garçons, le taux de diplômés de l'enseignement supérieur se stabilisant autour de 18-19 %. Ces nouveaux diplômés continuent de privilégier de plus en plus les filières professionnalisantes courtes. Pour les filles de ces mêmes générations, l'augmentation du nombre de diplômées de l'enseignement supérieur se poursuit, dans le prolongement de la poursuite de la diffusion du baccalauréat. La part de diplômées du supérieur passe de 18 % à 22 %. Les matières littéraires connaissent une désaffection relative, compensée par un attrait grandissant pour les études de droit et d'économie. Pour les garçons comme pour les filles, la mise en place d'un *numerus clausus* pour les études de médecine en 1971 réduit nettement le poids de ces diplômés dans l'ensemble du supérieur.

### ***1.3 Générations 1960-1977 : l'ouverture du supérieur reprend***

#### **La diffusion s'accélère chez les filles et reprend pour les garçons**

La deuxième phase de forte croissance du supérieur concerne les générations nées entre 1960 et 1977 ; la diffusion s'accélère chez les filles et reprend pour les garçons<sup>1</sup>.

---

<sup>1</sup> Chauvel, travaillant sur l'âge médian de fin d'études, constate également que « le niveau monte (...) mais par vagues successives » ; il isole deux périodes : « une première pour les cohortes nées au début des

Cette forte progression est le prolongement de l'expansion rapide des études secondaires. Résultat d'une volonté politique affirmée<sup>2</sup>, en l'espace de 15 ans, entre les générations 1960-62 et 1975-77, le taux de bacheliers a augmenté de 30 points pour les deux sexes (Annexe 3 - A1). Le taux de poursuite dans l'enseignement supérieur de ces nouveaux bacheliers se maintenant aux alentours de 63-66 %, la part de diplômés du supérieur augmente de 20 points. En quinze ans, la part d'une génération qui obtient un diplôme supérieur au baccalauréat double.

### **Une structure par filière qui reste stable**

Pour les hommes, hormis le recul des études de médecine, la répartition par filière reste relativement stable. Le développement des études professionnalisantes courtes reste fort, mais son rythme suit désormais celui de l'ensemble du supérieur. Pour les femmes, on observe même un léger recul du poids de ces filières parmi les diplômées du supérieur. Ce relatif recul ne concerne en réalité que les diplômes professionnalisants des domaines de l'éducation, de la santé et du social, très importants chez elles (Annexe 3 - A4). En effet, le contingent de places dans les écoles paramédicales ne suit pas le rythme de croissance du reste du supérieur et de plus en plus de futurs enseignants du primaire ne présentent les écoles normales qu'après une licence, devançant largement la création des IUFM (1991). Ce report des instituteurs vers les études universitaires bénéficie d'abord aux filières littéraires et scientifiques, mais les études de droit-économie-gestion se développent également.

### **Une diffusion qui concerne également le 2<sup>ème</sup> cycle, mais qui est moins manifeste pour le 3<sup>ème</sup> cycle**

Lors de cette deuxième phase d'expansion, la diffusion du supérieur s'est prolongée au-delà du premier cycle, et ce, en dépit du développement des filières « courtes » professionnalisantes<sup>3</sup>. Ainsi, la part de personnes diplômées d'au moins un deuxième cycle a elle aussi fortement augmenté. Elle a doublé en un peu plus de 10 ans : alors que 9 % de la génération 1960-62, hommes et femmes confondus, terminaient leurs études avec un diplôme supérieur ou égal à la licence, c'est désormais le cas de 19,5 % des hommes et de 25 % des femmes. Les taux de poursuite d'études au-delà du premier cycle général n'ont pas fléchi dans le temps. Ils se sont maintenus constants et à un niveau élevé : autour de 90 %. L'ouverture a été moins marquée pour les troisièmes cycles. Pour les hommes, les chances d'être diplômé d'un troisième cycle sont passées de 7 % pour la cohorte 1960-62 à 10,3 % pour celle née entre 1972 et 1974. Pour les femmes, ces chances ont doublé sur la période, passant de 4,3 % à 8,8 %.

### **Cette ouverture signe un changement de statut des études supérieures**

En près de cinquante ans, le statut des études supérieures a donc complètement changé. Réservées à une minorité pour la génération née au début des années 1930, elles ne concernaient alors que 8 % des hommes et 5 % des femmes, elles accueillent 39 % des hommes et 47 % des femmes nés au milieu des années 1970. Cet essor n'a pas été continu au fil du temps : l'université, et l'enseignement supérieur

---

années quarante et une seconde pour celles venues au monde à l'orée des années soixante-dix » (Chauvel, 1998).

<sup>2</sup> La création du bac technologique en 1968, puis du bac professionnel en 1985 répond à un souci d'élargir le recrutement social du baccalauréat. En 1987, le « plan pour l'avenir de l'Education nationale » (ou plan Monory) définit des objectifs quantitatifs en termes notamment d'accès au baccalauréat (Esquieu, Poulet-Coulibando, 2002).

<sup>3</sup> Nous avons classé tous les diplômes des formations professionnalisantes courtes en diplômes de niveau 1<sup>er</sup> cycle (voir annexe 1).

dans son ensemble, ont connu deux périodes de très forte croissance de leurs effectifs. La première se situe dans les années 60 et concerne les générations nées entre 1930 et la fin des années 1940. L'enseignement supérieur s'ouvre alors, mais reste relativement sélectif : c'est encore moins d'un jeune sur cinq de la génération née entre 1948 et 1950 qui est diplômé du supérieur. La seconde phase d'expansion du supérieur a lieu dans les années 1980 et se prolonge jusqu'au milieu des années 1990 ; elle concerne les générations nées entre 1960 et 1977. À l'issue de cette phase<sup>4</sup>, c'est 43 % d'une classe d'âge qui sort du système éducatif avec un diplôme du supérieur. Cette période est donc celle d'une massification manifeste du supérieur.

### **Une ouverture quantitative partagée par tous les milieux sociaux**

Cette ouverture des niveaux les plus élevés du système éducatif a été partagée par tous, quelle que soit leur origine sociale. Ainsi, entre les générations nées au début des années 1930 et celles nées au milieu des années 1970, la part de diplômés du supérieur est passée de 38 % à 77 % pour les enfants de cadres et de 2 % à 25 % pour les enfants d'ouvriers<sup>5</sup> (Annexe 3 - A6). S'il est difficile à ce stade de comparer l'ampleur de ces évolutions selon les origines sociales, tant les niveaux de probabilité sont différents, on peut noter le destin original des enfants d'agriculteurs. Alors que leurs chances d'achever leurs études avec un diplôme du supérieur étaient, pour les générations du début des années 1930, égales à celles des enfants d'ouvriers, pour les dernières générations observées, celles nées entre 1975 et 1977, elles sont deux fois supérieures (50 % contre 25 %).

Malgré la diffusion généralisée de l'enseignement supérieur, d'importantes différences entre milieux sociaux demeurent. La part des diplômés du supérieur parmi les enfants d'ouvriers nés au milieu des années 1970 est inférieure à celle des descendants de cadres nés dans les années 1930. Dès les générations nées dans les années 1940, être diplômé de l'enseignement supérieur constituait le destin scolaire majoritaire des enfants de cadres ou d'enseignants ; pour la génération 1975-77, seule une minorité des enfants d'ouvriers atteignent ce niveau. Les différences sociales apparaissent très marquées et peu progressives : plus de 20 points séparent les enfants de cadres des enfants dont les parents exercent une profession intermédiaire. Pour les filières professionnalisantes, la hiérarchie sociale reste respectée, mais est plus graduelle. Pour la génération 1975-77, la part de diplômés des filières professionnalisantes du supérieur s'échelonne entre 14 % pour les enfants d'ouvriers et 26 % pour les jeunes dont les parents exercent une profession intermédiaire (Annexe 3 - A7).

C'est sur la deuxième phase d'expansion du supérieur que nous concentrerons notre analyse. Lors de la première phase, l'université avait « absorbé » les nouveaux arrivants sans que la croissance de ses effectifs ne conduise à de profonds changements de son organisation alors que la seconde phase d'expansion a été de pair avec une mutation plus grande du paysage universitaire français. Surtout, lors de la première phase d'expansion, l'enseignement supérieur restait encore un univers relativement sélectif : les 20 % qui en sortaient diplômés pouvaient espérer que leur diplôme suffirait à leur garantir l'accès au statut de cadre. Lorsque 40 % d'une classe d'âge sort du système éducatif diplômée du supérieur, il est probable que les diplômes n'ont plus la même valeur et que l'importance du choix de l'orientation au sein du supérieur soit renforcée.

---

<sup>4</sup> La dernière génération observée, celle née entre 1975 et 1977 semble bien marquer la fin d'une période, celle de l'allongement de la durée des études. Ainsi, Durier et Poulet-Coulinbando (2004) indiquent qu'après avoir fortement augmenté, notamment autour de 1990, l'âge espéré de fin d'études est stable depuis le milieu des années 1990. En particulier, le développement massif connu par l'enseignement supérieur entre 1987 et 1995 semble depuis s'être essoufflé.

<sup>5</sup> Cela tient en partie à une transformation du milieu ouvrier sur la période (hausse du poids des ouvriers qualifiés), mais pas seulement : même au sein des seuls qualifiés ou non qualifiés, la progression a été manifeste.

## II - Une démocratisation à tous niveaux ?

### *II.1 Le débat autour de la question de la démocratisation de l'école*

Il est remarquable que la diffusion rapide du baccalauréat dans les générations récentes ait été de pair avec un maintien, quelle que soit l'origine sociale, de la propension à continuer des études supérieures (Annexe 3 - A10)<sup>6</sup>. On aurait pu penser que l'ouverture du bac à des enfants de milieux pour lesquels il représentait déjà une forte élévation du niveau d'instruction ne se répercuterait pas entièrement sur les effectifs du supérieur et que ces enfants, une fois leur bac en poche, et dotés d'un niveau d'instruction déjà bien supérieur à celui de leurs parents, arrêteraient leurs études. Cela n'a pas été le cas. L'enseignement supérieur a massivement accueilli à partir des années 80 un public qui en était jusque là largement exclu et ceci constitue en soi un progrès social incontestable. Mais l'ouverture de l'école ne signifie pas que les inégalités se sont réduites. À l'inverse, cette ouverture les rend plus sensibles : il pouvait être aisé de trouver des explications aux inégalités scolaires dans un système éducatif où les classes populaires et les classes bourgeoises ne fréquentaient pas la même école et où certains milieux sociaux étaient quasiment exclus de l'école au-delà d'un certain niveau scolaire ; dans un système unifié et où l'école s'est largement ouverte aux milieux défavorisés, la persistance d'inégalités sociales fortes devient embarrassante.

Le débat autour de la démocratisation de l'école reflète cette sensibilité croissante à cette question et est, au sein de la communauté scientifique aussi, relativement passionné. Il s'est longtemps cristallisé autour de questions de mesure<sup>7</sup>. Depuis que l'utilisation d'une échelle logistiqu (ou de odds ratios) fait à peu près consensus (Encadré 1), la plupart des études montrent que les proportions d'enfants arrivant à chacun des stades du cursus scolaire ont tendance à se rapprocher dans le temps. C'est le cas si on place à l'entrée de la 6<sup>ème</sup>, à l'entrée en seconde, au niveau du baccalauréat (Duru-Bellat, Kieffer, 2000), ou encore au niveau de n'importe quel cycle du supérieur (Selz, Vallet, 2006). Les structures par diplôme des enfants de différents milieux ont aussi tendance à s'homogénéiser au fil des générations (Thélot, Vallet, 2000). Au-delà de cette conclusion qui montre que les écarts de chances scolaires entre milieux se réduisent, deux résultats viennent la tempérer. Le premier est qu'il apparaît dans ces études que les grandes heures de la démocratisation concernent les générations nées de la fin des années 30 au début des années 50. Pour les générations suivantes, la démocratisation continue, mais à un rythme ralenti, alors même que ce sont ces générations qui ont été concernées par les réformes éducatives visant à réduire les inégalités (notamment la création d'un système scolaire unifié qui devait précisément être l'instrument d'un accès plus égalitaire aux certifications scolaires). Le deuxième constat de ces études est que la très large élévation du niveau scolaire s'est faite avec des inégalités sociales, certes moins prégnantes, mais toujours très élevées.

---

<sup>6</sup> D'autant plus que les nouvelles filières du baccalauréat, technologique et professionnelle, visent avant tout une insertion professionnelle immédiate et donnent moins souvent lieu à une poursuite d'études dans le supérieur (Esquieu, Poulet-Coulibando, 2002).

<sup>7</sup> On peut citer à ce propos le débat méthodologique qui a eu lieu il y a vingt ans dans la *Revue Française de Sociologie*, en réponse notamment à l'article de Combessie « L'évolution comparée des inégalités : problèmes statistiques » (1984) ; débat alimenté par les contributions de Grémy (1984), Florens (1984), Prévot (1984), Merlié (1985) et quelques années plus tard Vallet (1988).

### Encadré 1 : l'échelle logistique

Les années 1980 ont été animées par un vif débat sur la mesure appropriée des inégalités sociales de réussite scolaire, débat illustré par une série d'articles parus dans la *Revue Française de Sociologie*. Il existe deux façons simples de mesurer la distance entre deux nombres : leur rapport (ou écart relatif, échelle multiplicative) et leur différence (ou écart absolu, échelle additive). Le débat sur l'évolution des inégalités scolaires était d'autant plus vif que les résultats divergeaient selon la mesure des inégalités adoptée. Mesurés par la *différence* entre probabilités de devenir bachelier, les contrastes sociaux entre enfants de milieu cadre et enfants d'ouvriers s'atténuaient ; mesurées par le *rapport* des probabilités, ces inégalités avaient tendance à croître.

Ce débat a trouvé son dénouement autour du consensus que la manière appropriée de mesurer la distance entre deux probabilités n'était ni leur différence, ni leur rapport, mais leur contraste logistique (même si le débat resurgit régulièrement (Combessie, 2004 ; Morel, 2003). L'échelle logistique revient à raisonner *additivement* non pas sur la probabilité, comme le fait l'échelle additive, mais *sur une transformée non linéaire de cette probabilité* :

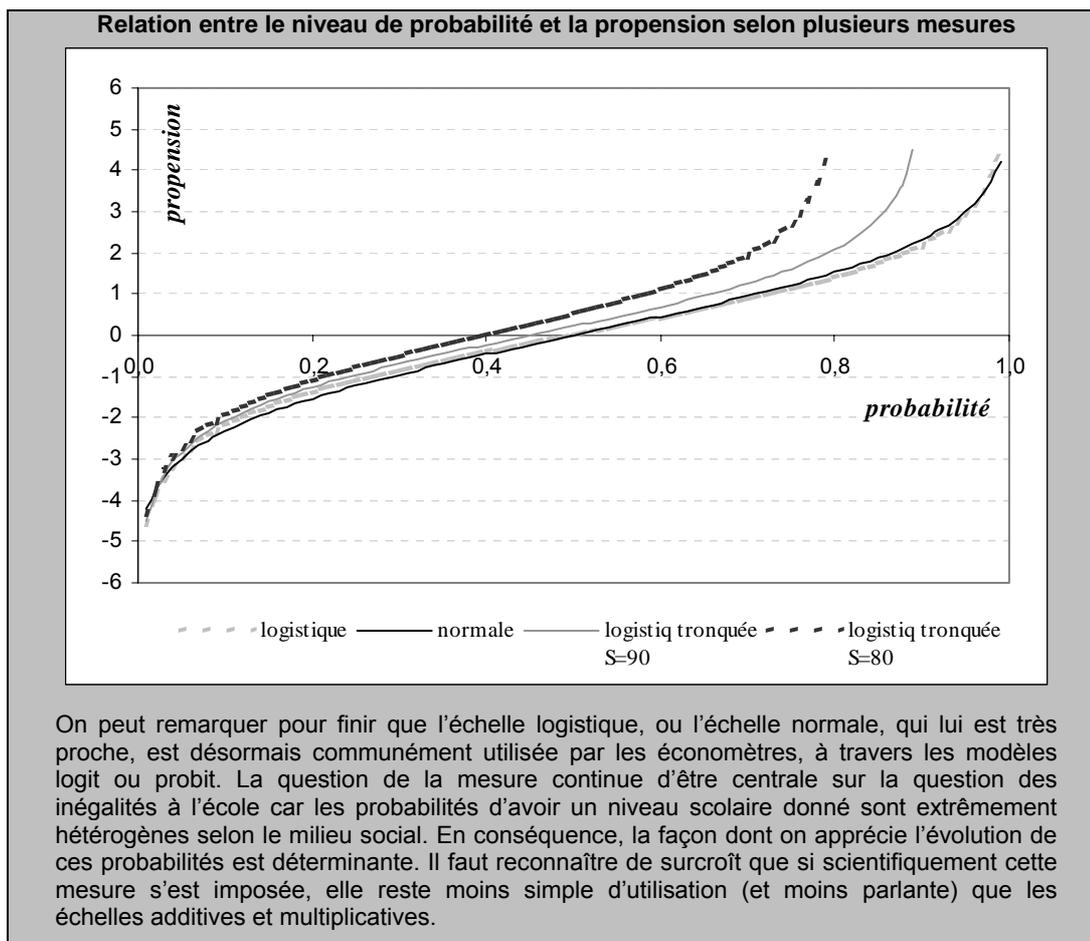
$$\ln\left(\frac{p}{1-p}\right)$$

Cette transformée de la probabilité est appelée propension. La distance entre deux probabilités, mettons 10 % et 15 %, est mesurée par la différence entre leurs propensions ; en l'occurrence, cette distance vaut 0,46 ( $\ln(0,15/0,85) - \ln(0,10/0,90)$ ).

L'idée intuitive sous jacente à ce choix de mesure est que l'appréciation de l'écart entre deux probabilités n'est pas indépendante du niveau de ces probabilités. Il est ainsi plus « facile » de passer de 40 % à 45 % que de passer de 1 % à 6 %. Cela disqualifie l'échelle additive (car  $45-40=5$  et  $6-1=5$ ). D'un autre côté, l'échelle multiplicative ne correspond pas forcément non plus à l'échelle de mesure intuitive que chacun a en tête. La plupart des gens vont considérer qu'il est plus facile de passer de 3 % à 6 % que de passer de 45 % à 90 % (et qu'il a donc plus de distance entre 45 % et 90 % qu'entre 3 % et 6 %), or l'échelle multiplicative dirait que l'écart est le même ( $6/3=90/45=2$ ).

Les statisticiens se sont progressivement accordés pour dire que l'on devait prendre en compte le niveau de la probabilité pour mesurer correctement son évolution. De même, ils s'accordent pour dire qu'il faut prendre en compte le caractère borné d'une probabilité (la marge d'accroissement d'une probabilité de 95 % est forcément limitée). L'échelle logistique répond à ces deux attentes. Cette mesure s'est d'autant plus facilement imposée dans les sciences sociales qu'elle correspond à plusieurs phénomènes réels. Elle s'est ainsi imposée en épidémiologie parce qu'elle correspond à la vitesse de diffusion d'une maladie contagieuse dans une population. Imaginons des personnes infectées d'un virus placées au sein d'une population non infectée et que chaque personne que croise une personne infectée au sein de la population est à son tour infectée ; la proportion de personnes infectées va ainsi suivre une courbe en  $\ln(p/1-p)$  (à un facteur près correspondant à la proportion initiale de personnes infectées). De même, la fonction logistique correspond au rythme de diffusion de biens nouveaux au sein de la population : au cours du temps, la proportion de personnes équipées du bien nouveau est d'abord très faible, puis augmente plus vivement avant de n'augmenter à nouveau modérément lorsqu'on s'approche du niveau de saturation dans la population.

On peut par ailleurs montrer que toute échelle de mesure sur des probabilités doit grosso modo avoir la forme d'une fonction logistique : être croissante et avoir un point d'inflexion. Cependant, alors qu'en épidémiologie ou dans le cas de diffusion d'un bien, l'échelle logistique correspond au processus de diffusion observé, dans le cas de l'école, rien n'impose de choisir précisément cette échelle. On peut choisir d'autres transformées des probabilités pour définir les propensions (pour peu qu'elles soient croissantes et aient un point d'inflexion). On peut ainsi préférer à la transformée logistique une échelle normale, dont la pente diffère aux deux extrémités de l'échelle logistique. On peut aussi considérer que le point de saturation n'est pas 100 % mais lui est inférieur et choisir une fonction d'expression  $\ln(p/S-p)$ , avec S, point de saturation, compris entre 0 et 1 (voir figure ci-dessous).



## ***II.2 Comparaison de la vitesse de démocratisation des différents cycles***

Au-delà de ce que l'on retient de ces résultats, baisse des inégalités ou maintien à un niveau élevé, ces études raisonnant à un niveau scolaire donné montrent que les écarts de chances entre les milieux diminuent. Plusieurs auteurs ont néanmoins souligné que cette mesure de la démocratisation de l'école n'était pas suffisante. Ils avancent l'hypothèse que face à la « banalisation » des diplômes du supérieur, notamment ceux du premier cycle, les enfants des milieux favorisés se seraient « repliés » sur les cycles longs, notamment les 3èmes cycles, aux rendements plus intéressants (Martinelli, Sigot, Vergniès, 1997 ; Givord, 2005). L'ouverture de l'école n'aurait fait que différer le moment où le gros de la sélection scolaire, et de la sélection sociale afférente, aurait lieu. Un tel schéma ne serait pas incompatible avec le fait que l'on mesure une démocratisation de l'obtention d'un diplôme de deuxième ou même de troisième cycle (ce qu'effectivement, on observe). Mais, sous cette hypothèse du « repli », la démocratisation des cycles longs devrait être moindre que celle mesurée au niveau des premiers cycles ou du baccalauréat. Nous testons la validité de cette hypothèse en comparant l'ampleur de la démocratisation à différents niveaux du supérieur. Deux possibilités se présentent naturellement pour comparer la démocratisation au niveau  $n-1$  et  $n$  : raisonner de façon conditionnelle et s'intéresser à la transition entre le niveau  $n-1$  et le niveau  $n$ , dans l'idée que ce qui se passe entre les deux niveaux nous informe sur la situation relative face à la démocratisation de ces deux niveaux, ou raisonner de façon non conditionnelle et comparer directement l'ampleur de la démocratisation aux deux niveaux scolaires.

## II.2.1 Comment comparer la démocratisation à deux niveaux de scolarité différents ?

La première possibilité pour comparer la démocratisation entre deux niveaux est de raisonner de façon conditionnelle. Des travaux récents ayant adopté simultanément une approche conditionnelle et non conditionnelle des inégalités sociales face au supérieur mettent en évidence un relâchement, désormais connu, du lien entre origine sociale et obtention d'un diplôme donné (même du supérieur), mais, fait moins connu, ils montrent également que les probabilités de sortir du système éducatif avec un diplôme du supérieur *conditionnellement au fait d'avoir le baccalauréat* sont plus disparates socialement aujourd'hui qu'hier (Figure 4 in Selz, Vallet, 2006.)<sup>8</sup>. Pouvons-nous en conclure, chose que ne font pas les auteurs, que l'enseignement supérieur s'est moins démocratisé que l'enseignement secondaire ? Pas nécessairement car la mesure utilisée pour apprécier la distance entre des probabilités, à savoir l'échelle logistique (encadré 1), malgré ses nombreuses propriétés intéressantes, présente la particularité de ne pas être transitive (car elle n'est pas linéaire). On peut ainsi mettre en évidence un paradoxe encore plus frappant où on a un *renforcement* des inégalités au niveau scolaire  $n-1$ , un *renforcement* des inégalités entre les niveaux  $n-1$  et  $n$ , mais globalement une *démocratisation* du niveau scolaire  $n$ . Le tableau suivant fournit un exemple numérique (et probable) où cela se produit.

	P(n-1)		P(n / n-1)		P(n)				
	t <sub>0</sub>	t <sub>1</sub>	t <sub>0</sub>	t <sub>1</sub>	t <sub>0</sub>	t <sub>1</sub>			
m <sub>1</sub>	61%	68%	45%	47%	27,5%	32%			
m <sub>2</sub>	20%	25%	25%	25%	5%	6,3%			
Odds ratio	6,3	<	6,4	2,5	<	2,7	7,2	>	7,0

On ne peut donc pas découper en séquences les parcours scolaires, comme le préconisait Mare (1980, 1981). Étudier séparément l'évolution des inégalités de chaque transition ne permet pas d'en déduire l'évolution des inégalités de l'ensemble du système scolaire. Appliqué aux résultats sur l'enseignement supérieur, on peut dire que les inégalités sociales des parcours scolaires parmi les bacheliers (probabilités conditionnelles d'avoir un diplôme du supérieur sachant qu'on a le baccalauréat) ont eu tendance à s'accroître, mais on ne peut rien en déduire sur l'évolution comparée des inégalités sociales au niveau du baccalauréat ou de l'enseignement supérieur. En particulier, on ne peut pas en conclure que le supérieur s'est moins démocratisé que le baccalauréat. La contradiction apparente entre les résultats issus des analyses conditionnelles et non conditionnelles vient de la propriété même de la fonction logistique qui apprécie les variations des probabilités en fonction de leur niveau initial.

La deuxième possibilité pour comparer la démocratisation entre deux niveaux est de raisonner de façon non conditionnelle. Mais la comparaison n'est pas immédiate ; elle est compliquée par le fait que les inégalités de départ peuvent être d'ampleur différente selon le point du cursus scolaire auquel on se place. Or, il est difficile de comparer l'évolution de odds ratio ayant des valeurs initiales très différentes.

Pour contourner ces difficultés, nous proposons une méthode consistant à isoler la démocratisation produite à un niveau scolaire  $n$  par la seule démocratisation des niveaux scolaires antérieurs. Nous construisons pour ce faire le contrefactuel suivant : les probabilités de passage du niveau  $n-1$  au niveau  $n$  sont restées inchangées dans le temps, pour chaque origine sociale. Sur la base de ces probabilités de transition fixes, nous calculons des probabilités fictives d'avoir un diplôme de niveau  $n$ . Les probabilités de transition variant dans le temps, mais modérément, les probabilités fictives et observées se situent à des niveaux de probabilité comparables et

<sup>8</sup> Le même type de résultat a été mis en évidence par Duru-Bellat et Kieffer (2000) à propos du collège : entre les générations 1949-53 et 1954-58, les inégalités entre cadres et ouvriers ont diminué pour l'accès en sixième et en seconde, mais les inégalités d'entrer en seconde parmi les élèves entrés en sixième ont augmenté.

correspondent à des structures d'inégalités proches. Concrètement, pour répondre à la question de savoir si le premier cycle s'est plus ou moins démocratisé que le deuxième, nous commençons par déterminer quel est l'impact de la démocratisation des premiers cycles sur les probabilités d'accès au deuxième cycle, en simulant quels auraient été les taux de diplômés pour chaque origine sociale si seule la sélection en amont du premier cycle s'était modifiée dans le temps (autrement dit si les taux de sélection entre le premier et le deuxième cycle avaient été constants). Nous comparons ensuite cette situation fictive à la réalité.

Cela peut être formalisé de la façon suivante. Construisons la probabilité simulée  $\hat{\pi}(2^e c / i, g)$  d'avoir un diplôme de 2<sup>ème</sup> cycle pour une personne du milieu  $i$  à la génération  $g$  de la façon suivante :

$$\hat{\pi}(2^e c / i, g) = \pi(1^{er} c / i, g) * \pi(2^e c / 1^e c, i, g_0)$$

Elle correspond à la probabilité d'accès au 2<sup>ème</sup> cycle si seule la probabilité d'accès au 1<sup>er</sup> cycle avait varié dans le temps, mais celle d'avoir un diplôme de 2<sup>ème</sup> cycle parmi les diplômés du premier étant restée constante. Les probabilités simulées et les probabilités réellement observées coïncident pour la première génération :

$$\hat{\pi}(2^e c / i, g_0) = \pi(2^e c / i, g_0)$$

Notons  $\Delta y(1^e c, g_0)$  et  $\Delta y(1^e c, g_1)$  les contrastes des taux de diplômés du premier cycle universitaire entre deux groupes sociaux pour la génération  $g_0$  et pour la génération  $g_1$  et  $\Delta y(2^e c, g_0)$  et  $\Delta y(2^e c, g_1)$  les contrastes pour le taux de diplômés du deuxième cycle. Les contrastes logistiques issus des probabilités observées et simulées coïncident aussi pour la première génération  $\Delta \hat{y}(2^e c, g_0) = \Delta y(2^e c, g_0)$  mais divergent ensuite pour les générations suivantes.

La comparaison des contrastes réels et simulés permet en théorie de conclure sur l'ampleur respective de la démocratisation aux deux niveaux de diplômes. Néanmoins, cette comparaison directe est rendue difficile par le fait qu'il y a 7 milieux sociaux, et donc 21 contrastes à comparer. Nous recourons donc à une modélisation pour synthétiser les évolutions de l'ensemble des odds ratio (encadré 2). Le modèle que nous utilisons, appelé modèle d'évolution uniforme des chances relatives dans la littérature sur la stratification sociale et proposé par Xie en 1992, décompose les contrastes logistiques selon l'équation :

$$\Delta y(2^e c, g_1) = \beta_{g_1} * \Delta y(2^e c, g_0)$$

soit, en évolution :

$$\Delta y(2^e c, g_1) - \Delta y(2^e c, g_0) = (\beta_{g_1} - 1) * \Delta y(2^e c, g_0)$$

Les contrastes à la date  $g_1$  dépendent de la structure des contrastes à la date initiale et d'un facteur temporel d'intensité.

En ajustant le modèle log multiplicatif de Xie sur ces probabilités simulées, on détermine de même  $\alpha_g$  tel que :

$$\Delta \hat{y}(2^e c, g_1) - \Delta \hat{y}(2^e c, g_0) = (\alpha_{g_1} - 1) * \Delta y(2^e c, g_0)$$

La comparaison des  $\alpha_{g_1}$  et  $\beta_{g_1}$  permet de comparer la démocratisation aux deux niveaux :

Si  $\alpha_{g_1} < \beta_{g_1}$ , cela signifie que la démocratisation des 1<sup>ers</sup> cycles ne suffit pas à expliquer celle qu'on observe au niveau des 2<sup>èmes</sup> cycles.

Si  $\alpha_{g_1} = \beta_{g_1}$ , cela signifie que la démocratisation des 2<sup>èmes</sup> cycles n'est que la conséquence de la démocratisation en amont.

Si  $\alpha_{g1} > \beta_{g1}$ , cela signifie qu'il y a eu un report de la sélection sociale : la démocratisation intervenue jusqu'au 1<sup>er</sup> cycle ne se traduit pas entièrement au niveau du 2<sup>ème</sup> cycle.

### Encadré 2 : Le modèle d'évolution uniforme des chances relatives

Pendant longtemps, l'étude des inégalités scolaires s'est centrée sur la comparaison des résultats scolaires des enfants de cadres et d'ouvriers. Ces deux milieux sont en effet emblématiques de toutes les oppositions sociales : oppositions économiques (riches versus moins riches), oppositions culturelles (familles dotées ou non d'un capital culturel proche de celui acquis à l'école). Mais surtout, élargir l'analyse à plus de deux milieux sociaux peut considérablement la compliquer. En ajoutant à l'analyse une seule origine sociale, les enfants d'employés par exemple, ce n'est déjà plus une évolution d'inégalités sociales qu'il faut suivre (différences de réussite scolaire entre enfants de cadres et d'ouvriers), mais trois (différences de réussite scolaire entre enfants de cadres et d'ouvriers, entre enfants de cadres et d'employés et enfin entre enfants d'employés et d'ouvriers). Rien n'assure que ces inégalités varient avec la même intensité, ni même dans le même sens : au cours du temps, les différences de réussite peuvent s'atténuer entre certains milieux et s'exacerber entre d'autres. Le risque d'avoir des évolutions divergentes selon les milieux considérés s'accroît quand on multiplie les milieux. Avec sept groupes sociaux, c'est 21 indicateurs d'inégalité dont on doit suivre l'évolution, chacun ayant sa dynamique propre.

La modélisation répond au souci de mieux prendre en compte la complexité de l'espace social (en considérant sept origines sociales plutôt que deux) tout en résumant les évolutions entre les différents milieux dans un scénario central unique plus intelligible.

À la date  $t$ , l'inégalité de réussite scolaire entre les enfants des milieux sociaux  $i$  et  $i'$  se mesure par le contraste logistique  $\Delta y_t$  (Encadré 1 et Annexe 4) :

$$\Delta y_t(i, i', j) = \log \left( \frac{p_{j|i}(t)}{1 - p_{j|i}(t)} \right) - \log \left( \frac{p_{j|i'}(t)}{1 - p_{j|i'}(t)} \right)$$

où  $p_{j|i}(t)$  est la probabilité pour une personne du milieu  $i$  née à la date  $t$  d'avoir le diplôme  $j$ .

Le modèle que nous utilisons a été développé par Xie (1992). Il ajuste sur le contraste logistique le modèle suivant :

$$\Delta y_t(i, i', j) = 2\beta_t(\psi_{ij} - \psi_{i'j}) \text{ sous la contrainte } \beta_{t_0} = 1.$$

Cette contrainte implique en particulier que  $\Delta y_{t_0}(i, i', j) = 2(\psi_{ij} - \psi_{i'j})$  (1)

Le contraste logistique  $\Delta y_t(i, i', j)$  dépend donc de la différence  $(\psi_{ij} - \psi_{i'j})$  qui traduit la forme du lien entre origine sociale et niveau de diplôme (elle est directement reliée aux contrastes sociaux de réussite à la date initiale d'après (1)) et du paramètre  $\beta_t$  qui exprime l'intensité de ce lien à la date  $t$ .

La propriété intéressante de ce modèle réside dans la forme particulière avec laquelle le milieu social interfère avec le facteur temporel. En effet, lorsqu'on regarde l'évolution du contraste logistique entre deux dates  $t_1$  et  $t_2$ , on obtient :

$$\Delta y_{t_2}(i, i', j) - \Delta y_{t_1}(i, i', j) = 2\beta_{t_2}(\psi_{ij} - \psi_{i'j}) - 2\beta_{t_1}(\psi_{ij} - \psi_{i'j}) = 2(\beta_{t_2} - \beta_{t_1})(\psi_{ij} - \psi_{i'j})$$

qui avec (1) donne :  $\Delta y_{t_2}(i, i', j) - \Delta y_{t_1}(i, i', j) = (\beta_{t_2} - \beta_{t_1})\Delta y_{t_0}(i, i', j)$

L'évolution des inégalités scolaires entre les milieux  $i$  et  $i'$  dépend ainsi d'un facteur temporel  $(\beta_{t_2} - \beta_{t_1})$  identique quels que soient les milieux comparés. L'évolution de l'ensemble des inégalités est ainsi résumée par la seule chronique des paramètres  $\beta_t$ .

Imaginons que l'on a 3 milieux sociaux ( $M_1, M_2, M_3$ ) ordonnés par réussite scolaire décroissante (les enfants issus de  $M_1$  réussissent mieux que ceux issus de  $M_2$  qui réussissent à leur tour mieux que ceux de  $M_3$ ) et que l'on regarde l'évolution des différences de réussite

scolaire à trois dates  $t_0, t_1, t_2$ . On a donc :

$$\forall i < i', \Delta y_{t_1}(M_i, M_{i'}, j) - \Delta y_{t_0}(M_i, M_{i'}, j) = (\beta_{t_1} - 1) \Delta y_{t_0}(M_i, M_{i'}, j)$$

$$\forall i < i', \Delta y_{t_2}(M_i, M_{i'}, j) - \Delta y_{t_1}(M_i, M_{i'}, j) = (\beta_{t_2} - \beta_{t_1}) \Delta y_{t_0}(M_i, M_{i'}, j)$$

Or du fait des caractéristiques des groupes  $M_i$  et  $M_{i'}$ , on a  $\Delta y_{t_0}(M_i, M_{i'}, j) > 0$ .

L'évolution des inégalités sociales par rapport à la date initiale nous est donc donnée par la comparaison à 1 de notre paramètre  $\beta_i$ . En particulier,  $\beta_{t_1} > 1$  implique

$$\Delta y_{t_1}(M_i, M_{i'}, j) > \Delta y_{t_0}(M_i, M_{i'}, j),$$

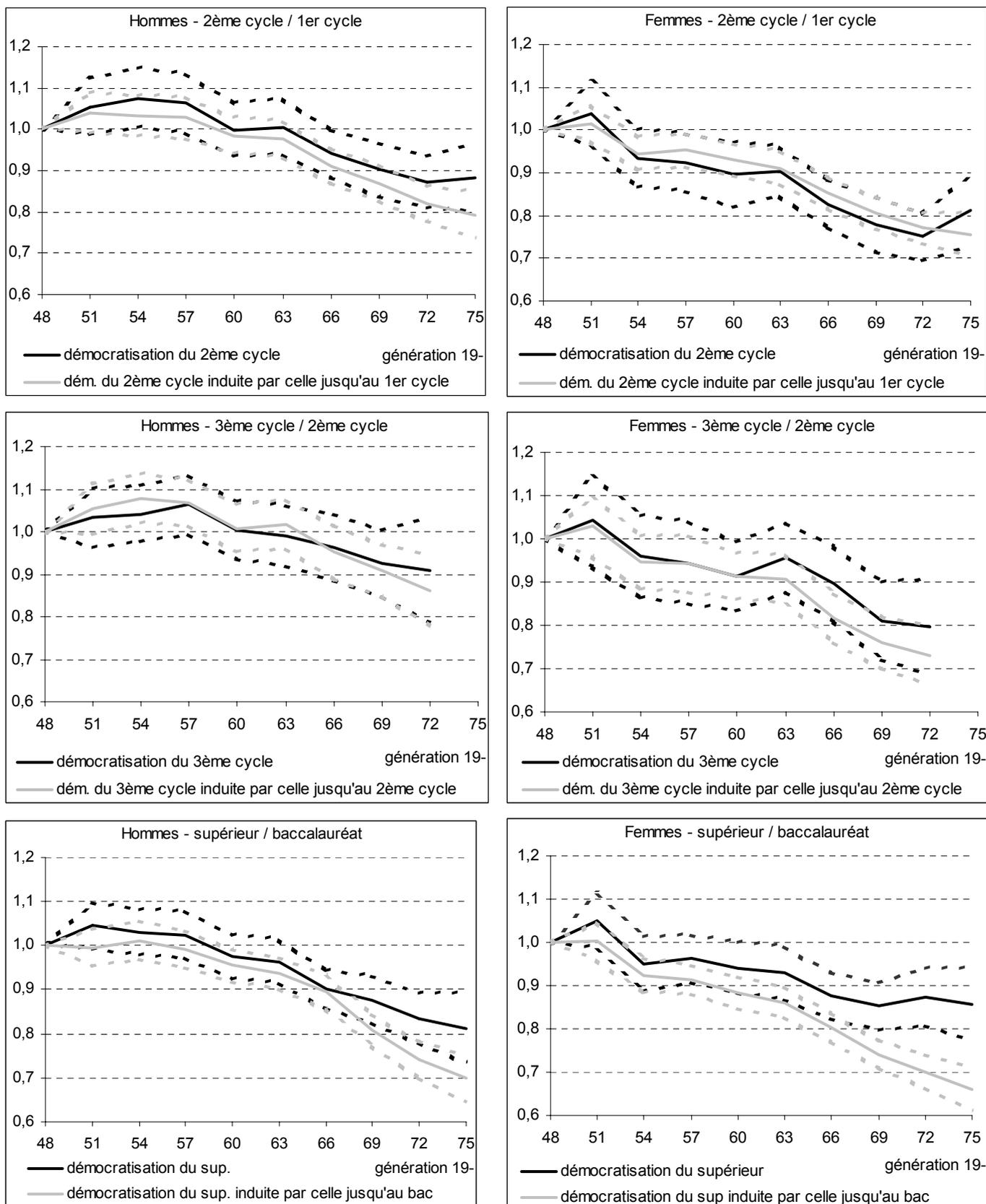
c'est-à-dire que les inégalités de réussite scolaire entre ces trois milieux sont plus fortes à la date  $t_1$  qu'elles ne l'étaient à la date  $t_0$ . Il convient de remarquer que cette évolution n'est pas celle connue pour deux milieux sociaux donnés, mais résulte de la tendance générale de nos trois origines sociales prises deux à deux. C'est pour cela que le modèle s'appelle modèle d'évolution *uniforme* des chances relatives. Rien n'empêche toutefois qu'à la date  $t_2$  les inégalités évoluent dans un sens différent. On peut très bien avoir  $\beta_{t_2} < \beta_{t_1}$ , ce qui signifierait que les inégalités ont eu tendance à diminuer entre  $t_1$  et  $t_2$ .

Le modèle de Xie bien qu'intéressant est aussi contraignant. Il suppose d'abord que sur l'ensemble de la période les différences de réussite scolaire entre deux milieux sont toujours dans le même sens (par exemple que les enfants de cadres réussissent toujours mieux que les enfants d'employés) ; cette propriété correspond toutefois à la réalité. Deuxièmement, il est pertinent si pour toutes les paires de milieux sociaux, les inégalités varient dans le même sens entre deux dates. C'est en général le cas, mais pas toujours.

## II.2.2 Les résultats

Les résultats ne confortent pas l'hypothèse d'une démocratisation au sein du supérieur qui s'essoufflerait au fur et à mesure que l'on progresserait dans le cursus. En effet, pour les hommes comme pour les femmes, la démocratisation a été de même ampleur au niveau du premier, du deuxième et du troisième cycle, ou, s'il existe des différences, elles ne sont pas significatives. Il ne semble pas qu'il existe un frein à la démocratisation à partir d'un certain niveau de diplôme, aussi élevé soit-il. En revanche, nos résultats montrent que l'enseignement supérieur, tous niveaux de diplôme confondus, s'est moins démocratisé que le baccalauréat. En effet, si les comportements de poursuite au-delà du bac étaient restés inchangés depuis la génération 1948-50, la démocratisation au niveau du supérieur aurait été supérieure à celle réellement observée. Une explication possible de la moindre démocratisation du supérieur par rapport à celle du baccalauréat est que l'ouverture du baccalauréat à des élèves qui auparavant n'accédaient pas à ce niveau tient en partie à la création des baccalauréats professionnels ; or leurs bacheliers poursuivent moins souvent des études que ceux issus des séries générales et technologiques (Esquieu, Poulet-Coulibando, 2002). En revanche, on aurait pu croire que le même type de mécanisme conduirait à ce que, au sein du supérieur, les premiers cycles se démocratisent plus que les deuxièmes cycles (les étudiants des filières professionnalisantes continuant rarement leurs études dans le deuxième cycle). Or nos résultats ne vont pas dans ce sens. Ceci s'explique par le fait que le poids des filières courtes professionnalisantes au sein du supérieur n'a pas autant augmenté que ce que l'on pourrait croire (il a même un peu baissé pour les femmes) ; le nombre de places dans ces filières a certes considérablement augmenté mais dans un contexte de forte croissance du nombre d'étudiants et d'une recomposition des filières professionnalisantes : le fort développement des IUT et STS a été compensé par un recul de la part des écoles paramédicales et sociales (Annexe 3 - graphiques A3 et A4).

**Graphiques 2 : Comparaison de la force du lien entre origine sociale et diplôme atteint pour les différents cycles du supérieur**



Note de lecture : Pour chacun des cycles, la courbe noire en trait plein représente l'ampleur de la démocratisation réellement observée. La courbe grise en trait plein représente la démocratisation que l'on aurait observée au niveau  $n$  si seule la démocratisation jusqu'au niveau  $n-1$  avait eu lieu. Plus la courbe descend, plus le lien entre origine sociale et diplôme se relâche et donc plus la démocratisation est forte. Les intervalles de confiance à 95 % sont en pointillé. Par exemple, pour le premier graphique en haut à gauche, comparant l'ampleur de la démocratisation dans le fait d'avoir un diplôme au moins de 1er cycle ou au moins de 2ème cycle, la courbe grise représente quelle aurait été l'ampleur de la démocratisation du 2ème cycle si seule la sélection sociale jusqu'au 1er cycle avait évolué, celle entre le 1er cycle et le 2ème cycle étant restée constante. Le fait que les deux courbes ne soient pas significativement différentes l'une de l'autre signifie que le rapprochement des probabilités entre les différentes origines sociales a été de même ampleur au niveau du 1er et du 2ème cycle.

Champ : le champ varie selon les graphiques, il correspond à chaque fois aux personnes dont on peut déterminer de façon certaine si elles ont obtenu ou non un diplôme au moins égal au cycle considéré.

### ***II.3 A taux de sélection constant, le supérieur s'est démocratisé***

L'idée selon laquelle avec l'expansion du supérieur ce sont surtout les cycles courts qui se sont démocratisés va souvent de pair avec celle que ces cycles courts, en se banalisant, ont perdu de leur valeur sur le marché du travail. En effet, la thèse du report suppose implicitement que le diplôme possède une valeur non pas absolue, mais relative ; l'un des rôles clés de l'école serait de classer les individus, et non pas de les former. Dans une situation où les capacités des individus seraient une information privée, inaccessible notamment aux employeurs éventuels, le diplôme serait le moyen pour les individus les plus capables de signaler leurs capacités. Dans cette lecture, l'objectif de l'éducation ne serait pas d'accroître la productivité des individus en augmentant leurs connaissances, mais de les classer selon leurs qualités naturelles (Arrow, 1973 ; Stiglitz, 1975). Il est certainement simpliste d'opposer ces deux lectures de l'apport de l'éducation, toutefois si la valeur du diplôme réside en partie dans le rang de classement qu'il confère aux diplômés, la diffusion des diplômes du supérieur n'a pu se faire sans une certaine dévalorisation de ses diplômes.

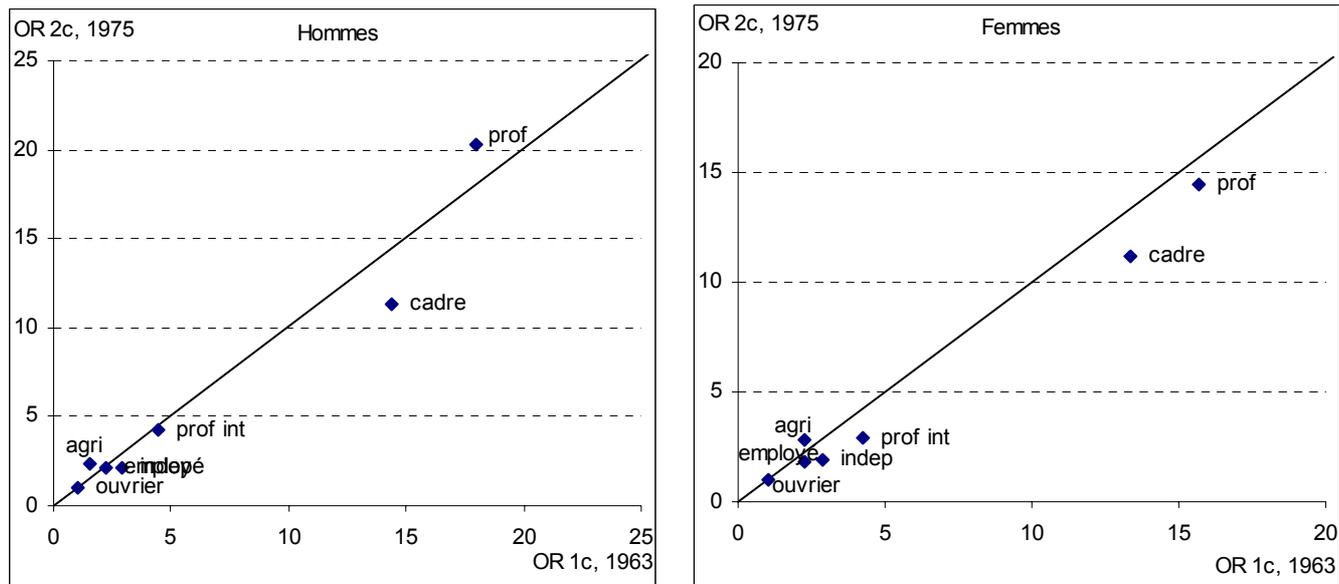
Pour prendre en compte cette dévalorisation des diplômes dans le temps, Thélot et Vallet (2000) examinent l'évolution de la structure par diplôme pour chaque milieu social en supposant que les diplômes se sont dévalorisés pour les générations nées après 1953. Ils apprécient cette dévaluation en appliquant de façon arbitraire une échelle d'équivalence entre les diplômes obtenus par les générations nées avant et après cette césure. Leur conclusion est inchangée : l'école s'est démocratisée, surtout pour les générations plus anciennes (nées dans les années 1930 et 1940). Goux et Maurin (1997) s'inscrivent encore plus franchement dans cette optique en raisonnant non plus en terme de niveau scolaire, mais en terme de classement. Ils retiennent en effet comme indicateur le rapport entre le nombre de fois où l'école respecte la hiérarchie sociale et le nombre de fois où elle l'inverse. Comparant la situation de la génération née entre 1935 et 1945 avec celle de la génération née entre 1959 et 1968, ils aboutissent à la conclusion que les inégalités n'ont que très peu diminué.

Une autre façon de répondre à la critique selon laquelle la démocratisation de l'école à un niveau scolaire donné ne signifie rien quand, dans le même temps, les diplômes se dévalorisent est de mesurer les inégalités des chances non pas à niveau de diplôme donné, mais à taux de sélection constant. L'idée est de se placer à deux dates  $t_1$  et  $t_2$  et de considérer deux niveaux  $n_1$  et  $n_2$  tels que la sélection globale qui s'exerce au niveau  $n_1$  en  $t_1$  est la même que celle qui s'exerce au niveau  $n_2$  en  $t_2$ . Ces deux niveaux peuvent alors être considérés comme des équivalents temporels du point de vue de la sélection qu'ils produisent. Cette situation s'est produite pour les 1<sup>er</sup> et 2<sup>ème</sup> cycles entre le début de la dernière phase de diffusion du supérieur et les dernières générations. En effet, 20 % de la génération née entre 1963 et 1965 avait au moins un diplôme du 1<sup>er</sup> cycle du supérieur ; la même part de la génération née entre 1975 et 1977, est sortie du système éducatif avec un diplôme de deuxième ou de troisième cycle du supérieur. À ces deux dates, les 1<sup>er</sup> et 2<sup>ème</sup> cycles présentent donc des taux de sélection semblables (20 %). Supposons que l'on scinde une génération en deux groupes : le groupe des « diplômés », qui regroupe les 20 % les plus diplômés, et le groupe des « moins diplômés » avec les 80 % restants ; pour la génération née entre 1963 et 1965, le groupe des « diplômés » est constitué des personnes ayant au moins un diplôme du 1<sup>er</sup> cycle du supérieur ; pour la génération née entre 1975 et 1977, par celles sorties du système éducatif avec un diplôme de deuxième ou de troisième cycle du supérieur. Dans le graphique représentant en abscisse les odds ratio d'être dans le groupe « diplômé » en  $t_1$  et en ordonnée ceux d'être dans le groupe « diplômé » en  $t_2$ . (avec comme référence les enfants d'ouvriers), les points illustrant le contraste avec les enfants de cadres sont systématiquement en dessous de la bissectrice (Graphique 3) : les chances relatives des enfants de cadres d'être dans le groupe des « diplômés » sont désormais plus faibles qu'elles ne l'étaient au début de la phase d'ouverture du supérieur. Ainsi, au

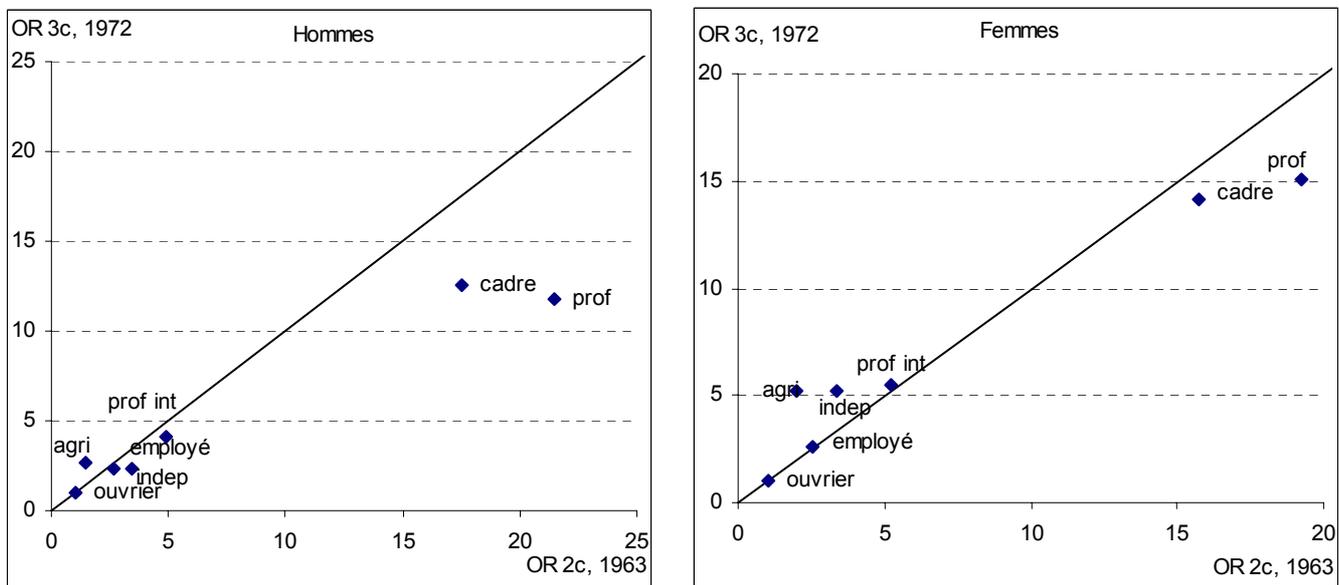
sein de la génération 1963-65, il y avait 13,6 fois plus de chances qu'un fils de cadre soit dans le groupe des diplômés et qu'un enfant d'ouvrier soit dans celui des moins diplômés que l'inverse. Pour la génération 1975-77, ce rapport n'est plus que de 8,5 (Graphique 3-a). On obtient les mêmes conclusions en considérant un taux de sélection plus strict, à 10 % (Graphique 3-b). Même en considérant qu'un diplôme ne vaut que par la place relative qu'il confère à son détenteur, on conclut que le supérieur se démocratise effectivement dans le temps.

**Graphiques 3 : Comparaison des inégalités sociales de certification scolaire à taux de sélection constant**

a) Sélection de 20 %



b) Sélection de 10 %



Lecture : pour la génération née entre 1963 et 1965, le odds ratio d'obtention d'un diplôme d'un niveau supérieur ou égal au 1<sup>er</sup> cycle entre fils de cadres et fils d'ouvriers était de 14,4. Le odds ratio d'obtention d'un diplôme de niveau supérieur ou égal au 2<sup>ème</sup> cycle pour la génération 1975-77 entre ces deux mêmes origines sociales était de 11,3.

### **III - Une polarisation sociale accrue des filières du supérieur ?**

Même si globalement l'enseignement supérieur s'est moins démocratisé que le bac, une appréhension plus fine de l'enseignement supérieur dans sa dimension « verticale » appuie plutôt l'idée d'une démocratisation manifeste et générale du supérieur, c'est-à-dire partagée par chacun de ses niveaux. Il convient à présent de se pencher sur l'évolution dans le temps des différences sociales dans « les choix d'éducation horizontaux ». Cette démocratisation globale de l'enseignement supérieur dissimule-t-elle une polarisation sociale accrue des différentes filières ? Autrement dit, toutes les filières ont-elles également participé à l'ouverture de l'enseignement supérieur ?

#### ***III.1 Les inégalités horizontales, une question ancienne relativement peu étudiée***

Cette question a été posée dans des travaux récents de Van de Werfhorst et Luijkx (2006) qui reviennent sur le constat d'une réduction des inégalités face à l'école aux Pays-Bas (Shavit, Blossfeld, 1993). Ces chercheurs valident empiriquement leur hypothèse selon laquelle le relâchement du lien entre origine sociale et position scolaire s'accompagne d'un maintien, voire d'un renforcement, des différences sociales dans les choix des domaines d'étude. C'est ce qu'ils nomment « Inequality Maintained through Horizontal Educational Choices ». En France, cette idée que les filières générales et professionnelles ont pour fonction de maintenir les cercles sociaux isolés et que, de cette façon, même dans les niveaux qui sont presque universellement atteints, des inégalités importantes peuvent se maintenir, est relativement ancienne. Dès 1986, Prost écrivait à propos de l'enseignement secondaire que « la création de filières nouvelles, notamment techniques, par des réformes qui prétendaient moderniser le système éducatif et œuvrer au développement technologique et économique, aurait permis de dériver une partie du flux montant des élèves d'origine populaire en direction des sections les moins prestigieuses, stabilisant ainsi leur place dans les filières dominantes du système, celles de l'enseignement général ». Pour le supérieur, Merle évoquait en 1996 un « modèle de développement différencié » selon lequel « l'ouverture de l'Université aux publics populaires [n'impliquerait] pas la démocratisation de l'élite scolaire dont le recrutement social demeurerait fermé ».

#### ***III.2 L'ouverture du supérieur s'est accompagnée d'une recomposition de ses filières***

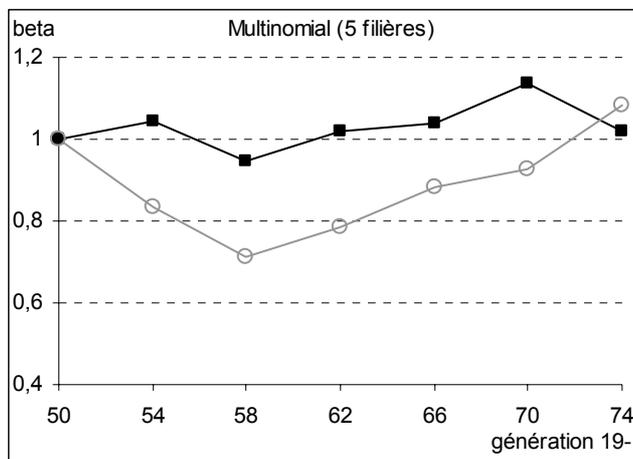
Il est vrai que l'ouverture du supérieur s'est accompagnée d'une recomposition de ses filières (Graphique 1), en raison d'un effet de demande (d'une part, les choix d'orientation des « nouveaux accédants » au supérieur ne sont pas forcément semblables à ceux des étudiants issus des milieux favorisés ; d'autre part, les choix de ces derniers ont pu eux-mêmes évoluer en réponse à l'ouverture du supérieur), mais aussi certainement en raison d'un effet d'offre (baisse du *numerus clausus* pour les études de santé et création des filières professionnalisantes IUT et STS par exemple). Les deux filières qui avaient un recrutement social marqué ont connu des évolutions contraires : les filières professionnalisantes, au recrutement plus populaire, se sont développées, tandis que les études de médecine davantage suivies par les étudiants issus des milieux favorisés se sont raréfiées. Ces évolutions ont-elles renforcé l'ancrage social de ces deux filières ? Plus généralement, le mouvement de démocratisation du supérieur est-il porté seulement par les filières qui étaient initialement socialement peu différenciées ?

### III.3 Les résultats

Nous testons cette hypothèse en comparant la polarisation sociale des différentes filières sur une période de faible expansion du supérieur (générations nées dans les années 1950) avec l'évolution de cette polarisation au moment où le supérieur s'ouvre à nouveau (générations nées à partir des années 1960). Ce faisant, nous cherchons à voir si l'éventuel renforcement (ou relâchement) de la polarisation sociale de certaines filières de l'enseignement supérieur peut être interprété comme résultant de stratégies socialement différenciées face à l'ouverture du supérieur.

Un modèle log-multiplicatif visant à étudier l'évolution dans le temps de l'association entre l'origine sociale et la filière d'enseignement, mesurée par une nomenclature en cinq catégories (professionnalisante, lettres et sciences humaines, sciences, droit-économie-gestion, médecine) met au jour des résultats contrastés selon le sexe (Graphique 4). Pour les hommes, il n'indique pas de tendance nette dans l'évolution de ce lien. Pour les femmes, il révèle qu'après s'être relâchées pour les générations nées dans les années 1950, les inégalités entre milieux sociaux dans le choix de la filière se sont ensuite renforcées jusqu'aux générations les plus récentes, à savoir celles nées dans le milieu des années 1970. Il convient toutefois de noter que ces évolutions, bien que significatives, restent modérées et que le lien entre origine sociale et filière, même s'il augmente pour les générations des années 1960 ne fait que retrouver son niveau initial. Cette évolution du lien donne un indice que, pour les filles au moins, s'est produit une « distillation fractionnée » (Prost, 1986) des nouveaux flux d'étudiants qui ont accédé au supérieur dans les années 1980. Mais la thèse de « Inequality Maintained through Horizontal Educational Choices » ou celle du « modèle de développement différencié », pour être testées, demandent à poursuivre l'analyse en considérant les filières prises isolément afin de voir si ce sont celles qui étaient initialement les plus sélectives socialement qui portent cette tendance.

**Graphique 4 : Évolution du lien entre filière (en 5 postes) et origine sociale pour les diplômés du supérieur**



Source : Enquêtes Emploi 1996 à 2002.

Note : Les courbes en noir avec des carrés concernent les garçons ; celles en gris avec des cercles les filles. Un paramètre qui décroît signifie que la polarisation des filières s'atténue.

Champ : personnes diplômées du supérieur.

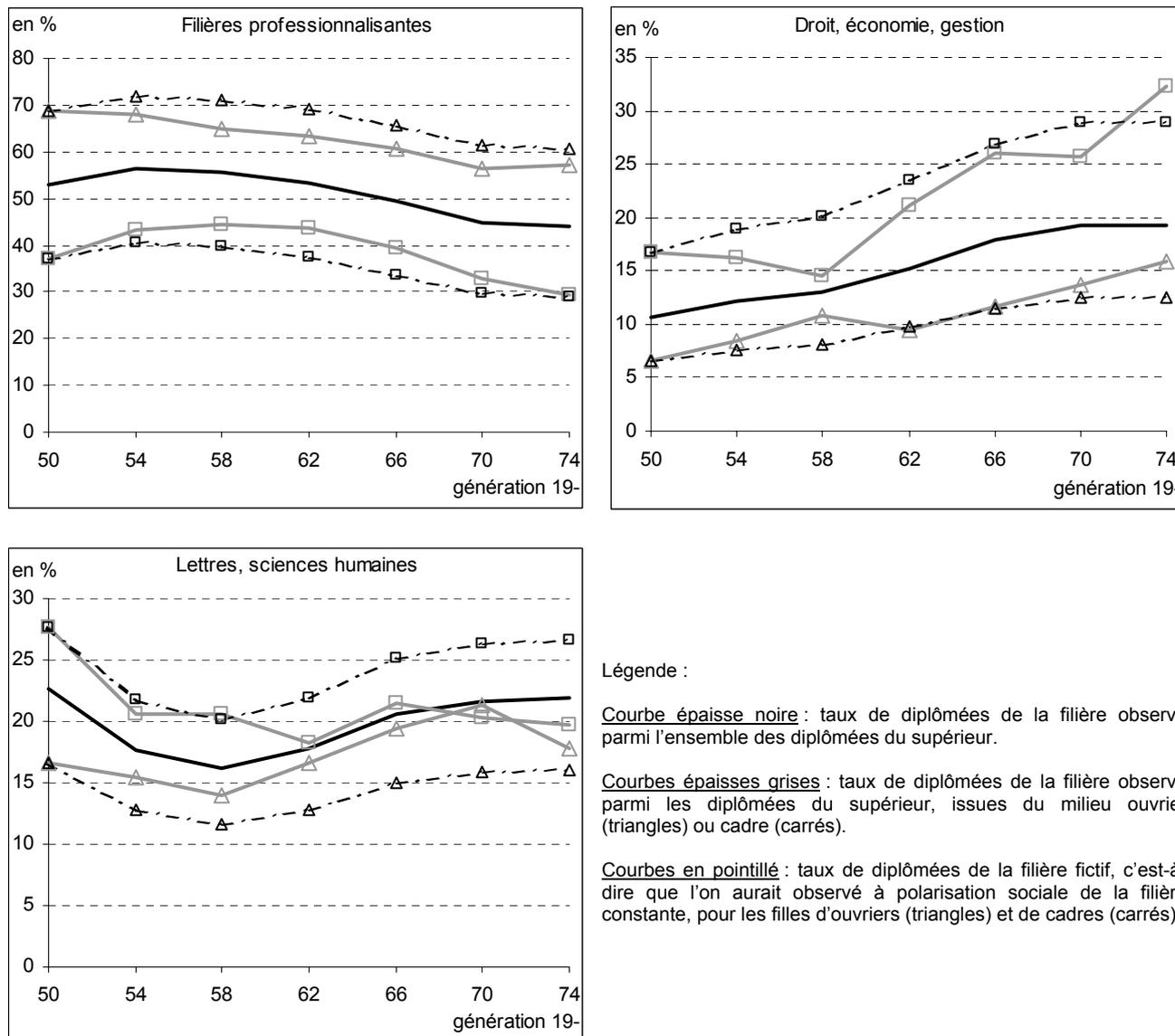
L'analyse détaillée confirme que pour les hommes, le lien entre origine sociale et spécialité se maintient dans le temps : au sein du supérieur, la démocratisation concerne également tous les domaines d'études (Annexe 5). Rappelons que le poids des diplômes des filières à visée professionnelle augmentait pour les garçons. Conformément à l'idée reçue, l'accès au supérieur des enfants des milieux les plus éloignés de la culture universitaire est largement passé par ces filières aux enseignements plus techniques et plus orientés vers l'acquisition de compétences professionnelles. Mais de manière moins attendue, le public de ces filières ne s'est pas « prolétarisé ». Les enfants de milieux plus favorisés se sont eux aussi tournés de

plus en plus vers ces filières, de telle manière que le gradient social est resté constant.

Pour les femmes en revanche, les filières dont le recrutement social était initialement le plus marqué, dans un sens ou dans un autre, (filières professionnalisantes, médecine et, dans une moindre mesure, droit-économie-gestion) se sont encore un peu plus polarisées socialement. Nous pouvons illustrer ce point en regardant le poids de l'enseignement professionnalisant au sein du supérieur pour deux milieux sociaux aux extrémités de l'éventail social : les enfants d'ouvriers et d'agriculteurs d'une part, et les enfants de cadres d'autre part. Pour les générations nées dans les années 1950, alors que les effectifs féminins des filières professionnalisantes évoluaient au même rythme que les effectifs de l'ensemble du supérieur, les filles de cadres ont opté de plus en plus souvent pour des études dans les filières professionnalisées, alors que les filles d'ouvriers qui faisaient des études supérieures se tournaient plus souvent vers l'université. La polarisation « populaire » de ces filières s'est donc atténuée. À partir des générations nées dans les années 1960, la situation change. Les filles d'ouvriers continuent à se tourner de plus en plus vers les études universitaires, au détriment des études professionnalisantes. Mais cette désaffection (relative) des filières professionnalisantes gagne aussi les filles de cadres, à un rythme encore plus soutenu, de sorte que relativement à l'ensemble du supérieur, la césure entre filières professionnalisantes, plutôt « populaires », et l'université, plus « bourgeoise », s'est accentuée. De la même manière, des générations 1950-53 à 1958-61, la part des études de droit-économie-gestion au sein du supérieur était légèrement en repli pour les enfants de cadres alors qu'elle progressait pour les enfants d'ouvriers. Le caractère « bourgeois » de ces études a donc commencé par s'atténuer. À partir de la génération 1962-64, l'attractivité de ces études a progressé chez les filles de cadres nettement plus vite que chez celles d'ouvriers, accentuant à nouveau la polarisation sociale de cette filière (Graphique 5). En revanche, la filière des disciplines littéraires, qui était initialement plus démocratique (le rapport des chances relatives entre les enfants de cadres et les enfants d'ouvriers avoisinait 2) a vu son recrutement social s'homogénéiser davantage encore. La convergence des taux de diplômées de cette filière parmi les diplômées du supérieur au sein de chaque milieu social d'origine est manifeste.

L'approche par filière permet donc d'étayer, pour les filles au moins, la thèse du développement différencié. La phase d'expansion rapide du supérieur, qui concerne les générations nées entre 1961 et 1976 a été de pair avec une accentuation de la polarisation sociale des différentes filières qui le composent : les filières les plus « bourgeoises » le sont devenues encore plus ; de même pour les filières les plus populaires. Avant cette phase d'expansion, lorsque le nombre de diplômées du supérieur croissait modérément, le mouvement était plutôt à l'atténuation de la polarisation sociale des filières du supérieur. Bien évidemment, la concordance ne suffit à établir une causalité, mais les faits concordent avec l'hypothèse du renforcement d'inégalités horizontales. Cette conclusion est toutefois largement mise en doute par le fait que cette augmentation des disparités sociales d'orientation ne se retrouve pas pour les garçons. L'expansion du supérieur les a pourtant tout autant concernés que les filles et il est difficile de comprendre pourquoi l'embourgeoisement du public féminin de filières très largement mixtes, comme la filière droit-économie-gestion, ne se retrouve pas pour le public masculin.

**Graphiques 5 : Évolution des probabilités d'être diplômée de certaines filières (conditionnellement au fait d'être diplômées du supérieur) pour les filles de cadres et d'ouvriers**



Source : enquêtes Emploi 1996 à 2002.

Champ : Femmes nées en France diplômées du supérieur âgées d'au moins 25 ans.

Note : Les probabilités fictives sont telles qu'elle maintienne constant le odds ratio entre une origine sociale donnée et l'ensemble des diplômées mesuré pour la génération 1949-1952.

Lecture : Pour la génération née entre 1950 et 1953, parmi les diplômées du supérieur issues du milieu ouvrier, 69 % étaient diplômées de la filière professionnalisante. Si les contrastes sociaux s'étaient maintenus constants, cette proportion aurait été de 72 % pour la génération 1954-57, alors qu'elle a été en réalité de 68 %.

## Bibliographie

Albouy V., Wanecq T. (2003), « Les inégalités sociales d'accès aux grandes écoles », *Économie et Statistique*, n° 361.

Albrecht J. W. (1975), « A procedure for testing the Signalling Hypothesis », *Journal of Public Economics*, vol. 15.

Arrow K. J. (1973), « Higher Education as a filter », *Journal of Public Economics*, vol. 2.

Combessie J. C. (1984), « L'évolution comparée des inégalités : problèmes statistiques », *Revue Française de Sociologie*, vol. XXV.

Combessie J. C. (2004), « Trente ans de comparaison des inégalités des chances : quand la méthode retenue conditionne la conclusion », *Courrier des statistiques*, n° 112.

Duru-Bellat M., Kieffer A. (2000), « La démocratisation de l'enseignement en France : polémiques autour d'une question d'actualité », *Population*, vol 55, n° 1.

Duru-Bellat M., Van Zanten A. (1999), *Sociologie de l'école*, Armand Colin.

Esquieu P., Poulet-Coulibando P. (2002), « Vers un enseignement secondaire de masse : 1985-2001 », in *Données Sociales - La société française*.

Florens J.-P. (1984), « Inégalité et dépendance statistique », *Revue Française de Sociologie*, vol. XXV.

Givord P. (2005), « Formes particulières d'emploi et insertion des jeunes », *Économie et Statistique*, n° 388-389.

Goux D., Maurin E. (1997), « Démocratisation de l'école et persistance des inégalités », *Économie et Statistique*, n° 306.

Grémy J.-P. (1984), « Sur les différences entre pourcentages et leur interprétation », *Revue Française de Sociologie*, vol. XXV.

Lemaire S. (2004), « Les bacheliers technologiques dans l'enseignement supérieur », *Éducation et formations*, n° 67.

Leridon H., Toulemon L. (1997), *Démographie : approche statistique et dynamique des populations*, Economica.

Mare R. D. (1980), « Social Background and School Continuation Decisions », *Journal of the American Statistical Association*, 75, pp. 295-305.

Mare R. D. (1981), « Change and Stability in Educational Stratification », *American Sociological Review*, 46, pp. 72-87.

Martinelli D., Sigot J.-C., Vergnies J.-F., (1997), « Diplômés de l'enseignement supérieur - l'insertion professionnelle se stabilise mais les écarts s'accroissent », *Céreq Bref*, n° 134.

Merle P. (1996), « Les transformations socio-démographiques des filières de l'enseignement supérieur de 1985 à 1995, essai d'interprétation », *Population*, n° 6.

Merle P. (2000), « Le concept de démocratisation de l'institution scolaire », *Population*, n° 1.

Merle P. (2002), « Démocratisation ou accroissement des inégalités scolaires ? : l'exemple de l'évolution de la durée des études en France (1988-1998) », *Population*, n° 4-5.

Merlié D. (1985), « Analyses de l'interaction entre variables. Problème statistique ou sociologique ? », *Revue Française de Sociologie*, vol. XXVI.

Ministère de l'Éducation nationale (2002), *Repères et références statistiques*.

Morel G. (2003), « Modéliser l'évolution des taux de scolarisation », *Mathématiques et sciences humaines*, 41<sup>ème</sup> année, n° 164, p 5-27.

Prévoit J. (1984), « A propos d'indices et de comparaisons de proportions », *Revue Française de Sociologie*, vol. XXV.

Prost A. (1986), *L'enseignement s'est-il démocratisé ?*, PUF.

Selz M., Vallet L.-A. (2006), « La démocratisation de l'enseignement et son paradoxe apparent », in *Données Sociales - La société française*.

Shavit Y., Arum R., Gamoran A. (2007), *Stratification in higher education: a comparative study*, Stanford University Press

Shavit Y., Blossfeld H.-P. (1993), *Persistent inequality. Changing educational attainment in thirteen countries*, Boulder, Westview Press.

Spence M. (1973), « Job Market Signaling », *The Quarterly Journal of Economics*, vol. 87, N°3.

Stiglitz J. E. (1975), « The Theory of "Screening", Education, and the Distribution of Income », *The American Economic Review*, Vol. 65, N°3.

Thélot C., Vallet L.-A. (2000), « La réduction des inégalités sociales devant l'école depuis le début du siècle », *Économie et Statistique*, n° 334.

Vallet L.-A. (1999), « Quarante années de mobilité sociale en France. L'évolution de la fluidité sociale à la lumière de modèles récents », *Revue française de sociologie*, 40(1), pp. 5-64.

Van de Werfhorst, Luijkx (2006), « Inequality Maintained through Horizontal Educational Choices, A Cohort Comparison for the Netherlands », working paper (<http://users.fmg.uva.nl/hvandewerfhorst/SocialSelectionInFieldsOfStudy.pdf> ).

Vermunt J.K. (1997), *LEM: A general program for the analysis of categorical data, Version 1.0*, Tilburg University.

Vermunt, J.K. (1996), *Log-linear event history analysis: a general approach with missing data, unobserved heterogeneity, and latent variables*, Tilburg University Press. Phd. thesis.

Xie Y. (1992), « The log-multiplicative layer effect for comparing mobility tables », *American Sociological Review*, 57.

## Annexe 1 : Données et nomenclatures

### **Données : l'enquête Emploi**

Entreprendre une analyse de l'évolution des chances devant l'enseignement supérieur nécessite non seulement des données homogènes et relativement fines sur le niveau de diplôme atteint et sur la spécialité de celui-ci, mais aussi de recouvrir une période longue qui prenne aussi en compte les générations les plus récentes, celles-là même qui ont été le plus concernées par le prolongement des études après le baccalauréat. D'un point de vue pratique, la principale difficulté d'une telle entreprise est de disposer d'un échantillon de taille suffisante pour pouvoir étudier en détail les différenciations au sein de l'enseignement supérieur. L'enquête Emploi, qui recueille annuellement auprès d'un large échantillon et avec un mode de collecte stable dans le temps des données relativement détaillées sur la formation des individus, semble constituer une source statistique adéquate.

Plus précisément, cette étude s'appuie sur une compilation des enquêtes Emploi annuelles réalisées entre 1990 et 2002, dernière collecte avant le passage aux enquêtes Emploi en continu<sup>9</sup>. L'échantillon des enquêtes Emploi étant renouvelé par tiers chaque année, seules les dernières interrogations individuelles ont été retenues. Nous nous assurons ainsi de ne pas compter plusieurs fois un même individu, tout en maximisant la taille de notre échantillon de travail. Pour les analyses par filière d'études, seules les enquêtes postérieures à 1996 peuvent être utilisées, les données relatives à la spécialité du diplôme étant jusqu'alors absentes ou trop lacunaires.

### **Le champ**

Dans chacune des enquêtes, ont été retenues les seules personnes nées entre 1930 et 1977, âgées d'au moins 25 ans, âge à partir duquel la part de personnes encore en études initiales et n'ayant toujours pas atteint le premier cycle du supérieur est négligeable (Annexe 2). Nous nous sommes en outre limitées aux personnes nées en France, afin de s'assurer au maximum que ces personnes ont effectué leurs études en France. On obtient au final un échantillon de 516 210 personnes dont 102 109 ont effectivement poursuivi des études supérieures<sup>10</sup>.

Bien que considérant l'ensemble de ces générations pour situer la diffusion de l'enseignement du supérieur sur longue période, nos analyses se concentrent par la suite sur les générations nées à partir des années 1950. Et ce pour plusieurs raisons. Tout d'abord, si l'explosion du supérieur concerne avant tout les générations nées après 1960, nous avons élargi la période d'étude aux générations antérieures afin de préciser si la période où les études supérieures se sont le plus fortement diffusées a été marquée par des évolutions spécifiques en terme de sélectivité sociale des filières ou des niveaux du supérieur. Il s'agit également des premières générations qui vont connaître l'enseignement supérieur tel qu'il a été redessiné par les différentes réformes des années 60. Elles obtiendront en effet leur baccalauréat à la fin des années 1960, soit juste après la création des IUT et des STS. Surtout, les

---

<sup>9</sup> Si l'utilisation des enquêtes Emploi en continu aurait permis de prolonger l'étude pour les générations les plus récentes, le choix a été fait de ne pas les inclure à l'analyse. En effet, le passage à l'enquête Emploi en continu a été l'occasion d'une refonte du questionnaire qui a notamment concerné la nomenclature des diplômes du supérieur (la modalité « certificat d'aptitude pédagogique, certificat d'études normales » (DIES=45), bien que concernant encore beaucoup de personnes dans les générations plus anciennes, n'existe plus) et la spécialité du diplôme (celle-ci n'est plus recueillie pour la formation en cours et est présentée sous une nomenclature légèrement différente à celle utilisée précédemment). Ces ruptures de série potentielles sont d'autant plus problématiques du fait du faible recouvrement avec les enquêtes emploi annuelles et du faible gain pour l'analyse (une génération supplémentaire).

<sup>10</sup> Et respectivement 497 834 et 99 674 si on exclut ceux dont la catégorie socioprofessionnelle du père n'est pas renseignée.

comparaisons temporelles des contrastes sociaux de réussite scolaire souffrent du fait que les contrastes entre milieux, contrastes économiques ou culturels, ne sont pas forcément constants sur longue période. On prend souvent l'exemple du groupe des exploitants agricoles qui s'est profondément renouvelé au cours du 20<sup>ème</sup> siècle. Dans le même temps, le statut de cadre, en même temps qu'il se diffusait, s'est banalisé. Il y a donc tout lieu de croire que les contrastes entre milieux se sont plutôt estompés et il peut être légitime de ne lire dans l'homogénéisation des niveaux scolaires des enfants que l'homogénéisation de leur environnement familial. Réduire la fenêtre temporelle, en se restreignant aux individus nés sur environ 25 ans, devrait limiter ces problèmes de comparabilité temporelle. Les contrastes entre milieux ne sont pas pour autant constants. Les contrastes économiques entre milieux ont eu par exemple tendance à légèrement s'amplifier (des années 50 au milieu des années 60) avant de s'atténuer jusqu'aux années 80 (revenant au niveau de contraste du début des années 50)<sup>11</sup>.

Les individus ont été regroupés, selon leur année de naissance, en des cohortes de trois ans pour l'étude par cycle d'enseignement et de quatre ans pour l'étude par filière.

### ***Les nomenclatures***

Les études sur la démocratisation supposent que soit définie une segmentation de l'espace social, l'objet étant de voir si cette segmentation devient au fil du temps de moins en moins pertinente pour rendre compte des différences de réussite scolaire. Cette segmentation peut se baser sur différentes dimensions : économique (revenu des parents), culturelle (diplôme des parents) ou socioprofessionnelle. Dans cette étude, comme dans de nombreuses autres, nous mesurerons l'origine sociale par la catégorie socioprofessionnelle (CS) du père, la seule recueillie dans l'enquête Emploi<sup>12</sup>. L'indicateur d'origine sociale retenu vise à rendre compte au mieux des différences de parcours au sein du supérieur. Il possède sept modalités : enseignants<sup>13</sup> ; cadres, professions libérales et chefs d'entreprise ; professions intermédiaires ; artisans et commerçants ; employés ; ouvriers ; agriculteurs.

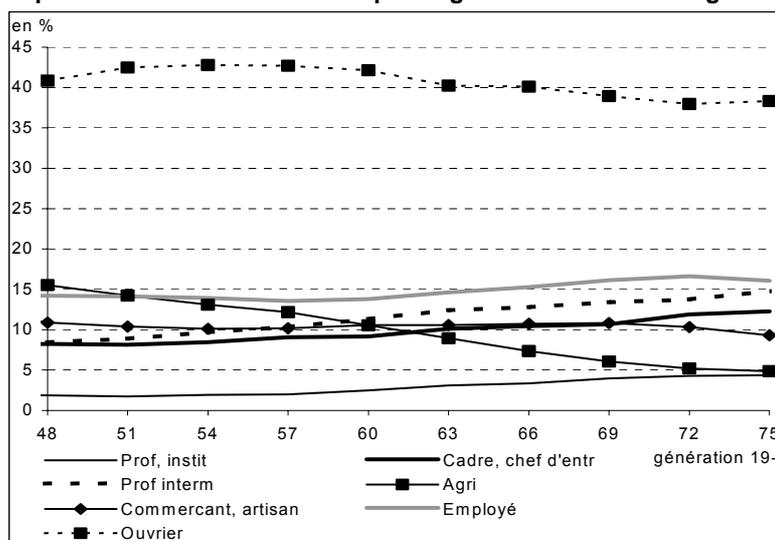
---

<sup>11</sup> Source : séries longues sur les salaires issues des DADS, INSEE.

<sup>12</sup> Thélot et Vallet (2000), exploitant les enquêtes Formation et Qualification Professionnelle, plus riches sur l'appréhension de l'origine sociale et culturelle des individus, montrent que leurs conclusions ne dépendent pas de la façon d'appréhender le milieu d'origine. Ils précisent toutefois que « les inégalités culturelles devant l'école se sont moins sensiblement réduites que les inégalités sociales ». Aussi, on peut supposer que si les inégalités sociales se renforcent, il devrait en être au moins de même pour les inégalités culturelles.

<sup>13</sup> Les enfants d'enseignants, bien que peu nombreux, sont isolés en raison de leur investissement scolaire atypique (Thélot, Vallet, 2000, Albouy, Wanecq, 2003).

Graphique 1 : Évolution de structure par origine sociale au fil des générations



Source : enquêtes Emploi, 1990-2002.

L'étude, à la suite d'autres travaux sur la démocratisation de l'école (Goux, Maurin, 1997, Thélot, Vallet, 2000), retient comme indicateur de la position scolaire acquise le plus haut diplôme obtenu et non le plus haut niveau atteint et ce pour deux raisons. D'une part, l'un et l'autre coïncident dans ce nombreux cas ; d'autre part, c'est avant tout la certification assurée par le diplôme qui peut être valorisée sur le marché du travail. Plus précisément, on distingue trois niveaux de diplôme au sein du supérieur, chacun correspondant aux différents cycles d'enseignement qui le composent. Ce découpage ne se superpose pas à une segmentation par type d'établissement fréquenté (université, grandes écoles, IUT...).

Notre étude se situe avant l'harmonisation européenne des cursus universitaires et la structuration sur le modèle licence/master/doctorat. Les générations étudiées ont donc évolué dans un enseignement supérieur qui délivrait des diplômes qui peuvent être regroupés en trois niveaux : premier, deuxième et troisième cycle. Le premier cycle regroupe les DEUG, les diplômes des professions de santé et du secteur social<sup>14</sup>, le certificat de fin d'études normales, les DUT et BTS et autres diplômes de techniciens. Le deuxième cycle d'enseignement correspond quant à lui aux diplômes de licence et maîtrise<sup>15</sup>. Le troisième cycle rassemble enfin les diplômes de niveau DEA, DESS, doctorat ou diplômes de grandes écoles<sup>16</sup>. Le tableau suivant précise la correspondance exacte entre le diplôme obtenu (variable DIES de l'enquête Emploi annuelle) et le cycle d'enseignement.

<sup>14</sup> Bien que certaines formations du secteur paramédical ou social durent trois ans.

<sup>15</sup> Bien que constituant un seuil important dans les transitions du supérieur selon P. Merle (2002), la licence et la maîtrise n'ont pas été distinguées, cette précision n'ayant été introduite dans l'enquête Emploi que depuis 1992.

<sup>16</sup> D'autres travaux se sont attachés spécifiquement à l'analyse des grandes écoles en les comparant en particulier aux autres diplômes de troisième cycle (Albouy, Wanecq, 2003).

**Tableau 1 : Correspondance entre les modalités de DIES  
(plus haut diplôme du supérieur obtenu) et les cycles d'enseignement**

Modalités de DIES	Cycle
41 - diplômes universitaires du 1 <sup>er</sup> cycle (sauf DUT et DEUST) : DEUG, DUEL, DUES, PCEM,...	1 <sup>er</sup> cycle général
42 - DUT (diplôme universitaire de technologie), BTS (brevet de technicien supérieur), DEUST...	1 <sup>er</sup> cycle professionnalisant
43 - diplômes de niveau de technicien supérieur (sauf ci-dessus) hormis dans les domaines de la santé et du secteur social : Brevet des Métiers d'Art, Diplôme National d'Arts Plastiques, Diplôme National d'Arts et Techniques, DPECF, clerc de notaire...	
44 - diplômes des professions de la santé et du secteur social sagefemme, infirmière, kinésithérapeute, laborantin, manipulateur radio-assistante sociale, éducateur spécialisé...	
45 - certificat d'aptitude pédagogique, certificat de fin d'études normales (CFEN), y compris diplôme de maître d'éducation physique	2 <sup>ème</sup> cycle
46 - diplômes universitaires du 2e cycle (licence, maîtrise)	
47 - diplômes universitaires du 3e cycle (DESS, magistères, DEA, doctorat, y compris doctorat en médecine, diplôme de chirurgien dentiste, CAPES, CAPET, agrégation)	3 <sup>ème</sup> cycle
48 - diplômes des écoles (non citées ci-dessous) d'ingénieurs, de commerce, d'études comptables supérieures, diplôme de notaire, d'avocat d'architecte, d'art, de journalisme, de vétérinaire...	
49 - diplôme de l'une des grandes écoles suivantes : Centrale, École de l'Air, Ec.de Magistrature, ESSEC, ENA, ENGREF, ENSAE 1ère div., Génie maritime, HEC, INA (Agro), Mines, Navale, Normale sup., Polytechnique, Ponts, Saint-Cyr, Sc. Po Paris, Sup-aéro, Télécom, Cons.nat.sup.art.	

Au sein de l'enseignement supérieur, on considère cinq filières construites comme le croisement entre le caractère professionnalisant ou généraliste des études et le domaine d'études. L'indicateur du domaine d'études résulte du regroupement de la nomenclature très détaillée, en 93 postes, de la spécialité du plus haut diplôme obtenu ou de celle de la formation actuellement suivie pour les personnes en cours d'études initiales. La première filière est constituée des formations professionnalisantes courtes, quel que soit leur domaine d'étude<sup>17</sup>. Ces formations ont pour dénominateur commun de dispenser un savoir axé sur l'acquisition de compétences professionnelles. Elles ne proposent qu'un nombre contingenté de places et sont accessibles le plus souvent sur dossier. Cet ensemble des formations professionnalisantes courtes rassemble les STS, les IUT, les écoles paramédicales et sociales, les écoles normales ainsi que les formations préparant à un diplôme de niveau technicien supérieur (comme le diplôme national d'arts plastiques). Les quatre autres filières sont constituées en fonction de la proximité de leur contenu disciplinaire. Elles regroupent des disciplines qui peuvent être dispensées aussi bien à l'université que dans des écoles (écoles de commerce, d'ingénieurs...). Au sein des études généralistes, les champs disciplinaires agrégés ainsi isolés sont : les lettres et les sciences humaines (langues, lettres, psychologie, histoire...), les sciences (mathématiques, sciences naturelles, physique, chimie...), l'économie, le droit et la gestion et la médecine (incluant les études de pharmacie, d'odontologie et de vétérinaires).

<sup>17</sup> Ponctuellement, nous isolerons en leur sein les études du secteur primaire et secondaire de celles du secteur tertiaire. Cette distinction n'est pas maintenue pour l'ensemble dans la mesure où les études professionnalisantes dans les domaines de l'agriculture et de l'industrie sont largement minoritaires pour les filles.

**Tableau 2 : Effectifs bruts des différentes filières et principales spécialités de ces filières**

Filières professionnalisantes	Droit, économie, gestion	Lettres, sciences humaines	Médecine	Sciences
<b>Hommes</b>				
10 312	4 185	2 264	1 175	6 044
Commerce, vente : 13,2%	Droit, sc. politique : 31,5%	Langues, civilisations : 14,0%		Physique-chimie : 13,5%
Électricité, électronique : 9,7%	Commerce, vente : 19,9%	Histoire : 13,4%		Informatique : 10,2%
Technologies indus. : 9,3%	Comptabilité gestion : 18,9%	Lettres : 11,8%		Électricité, électronique : 8,8%
Comptabilité, gestion : 8,8%	Économie : 12,4%	Disciplines artistiques : 12,2%		Génie civil : 7,6%
Santé : 7,4%		Sciences sociales : 7,8%		Mathématiques : 7,1%
Spécialités pluritechnologiques : 7,5%		Psychologie : 6,1%		
<b>Femmes</b>				
14 222	4 479	5 672	1 235	2 763
Santé : 31,0%	Droit, sc. po : 40,5%	Langues, civilisations : 29,9%		Sc. de la vie et de la terre : 21,2%
Secrétariat : 14,6%	Comptabilité, gestion : 15,0%	Lettres : 16,9%		Physique-chimie : 16,7%
Comptabilité, gestion : 10,4%	Commerce, vente : 14,8%	Psychologie : 12,3%		Mathématiques : 12,5%
Enseignement : 9,9%	Économie : 9,7%	Disciplines artistiques : 7,9%		
Commerce, vente : 9,4%	Spécialités plurivalentes des échanges et de la gestion : 9,6%	Sciences sociales : 7,8%		
		Histoire : 7,4%		

Champ : personnes diplômées du supérieur dont la spécialité du plus haut diplôme ou de la formation actuellement suivie est renseignée et nées entre 1950 et 1977.

Note : dans le détail des filières, seules les filières les plus importantes, classées par ordre décroissant sont citées.

Lecture : 10 312 hommes de notre champ ont un diplôme d'une filière professionnalisante du supérieur ; pour 13,2 % d'entre eux, il s'agit d'un diplôme dans le domaine du commerce ou de la vente et pour 9,7 % d'un diplôme en électricité ou électronique

## Annexe 2 : Quel champ retenir en termes d'âge ?

L'enquête Emploi recueille, entre autres, le plus haut diplôme obtenu dans l'enseignement supérieur, pour les personnes âgées de 15 ans ou plus vivant en ménage ordinaire. La compilation de différentes enquêtes Emploi (ici, les enquêtes Emploi annuelles de 1990 à 2002) permet de travailler sur un échantillon de taille importante et d'observer comment la structure des diplômes se déforme au fil des générations. Si le mode de questionnement est relativement constant dans le temps et autorise une telle entreprise, il convient de s'interroger sur les âges auxquels les différentes cohortes peuvent être observées. Cette question se pose tout particulièrement aux deux extrémités de la vie, mais dans des termes différents. Pour les plus jeunes, le fait qu'une part non négligeable d'entre eux soit encore en études initiales risque de sous-estimer leur niveau de diplôme atteint. Pour les plus âgés, on se heurte à des problèmes de déclaration plus importants, car portant sur un événement plus lointain, et à des biais de sélection liés au fait que, à un âge donné, ne sont observées que les personnes encore présentes (c'est-à-dire non décédées ou qui n'ont pas quitté le territoire), soit une fraction de la cohorte.

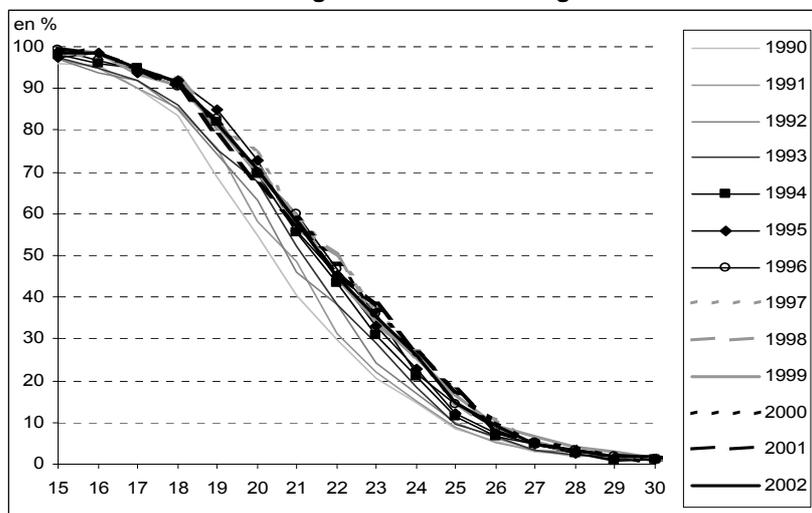
### Quelle borne inférieure en termes d'âge retenir ?

La première question est de savoir à partir de quel âge on peut inclure les personnes dans notre champ, autrement dit, quel est l'âge minimal pour lequel on connaît, ou peut connaître, le plus haut diplôme obtenu.

On peut considérer deux grandes catégories de personnes : celles qui ont fini leurs études initiales et pour lesquelles on connaît de façon certaine le plus haut diplôme obtenu et celles en cours d'études pour lesquelles on ne connaît que le plus haut diplôme atteint au jour de l'enquête et la formation suivie à ce moment-là.

Le premier point que l'on peut alors regarder est le poids de ces deux groupes au fil de l'âge, autrement dit, la part des personnes en études initiales selon l'âge (Graphique 1). Cette part décroît continûment au fil de l'âge. De plus, à un âge donné, la part de personnes encore en études augmente au fil des enquêtes, traduisant la progression de la scolarisation au fil des générations. Ce n'est donc qu'à partir d'un âge relativement avancé que nous disposons d'une information certaine sur le plus haut niveau de diplôme atteint, âge d'autant plus élevé que nous considérons des générations récentes.

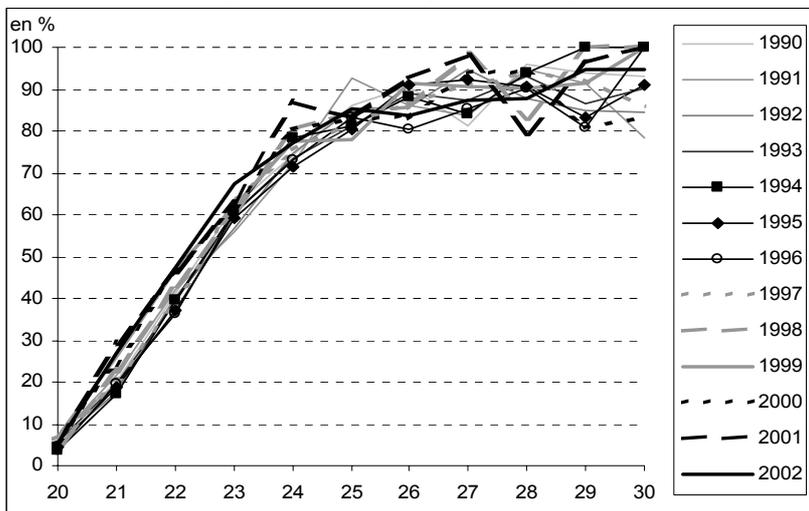
**Graphique 1 - Part de personnes encore en études initiales selon l'âge et la date d'interrogation**



Source : enquêtes Emploi 1990-2002, Insee.

Toutefois, parmi les personnes n'ayant pas encore achevé leurs études, certaines ont déjà atteint un certain niveau dans la hiérarchie scolaire, de sorte que nous disposons pour elles d'un niveau plancher sur leur plus haut diplôme atteint. Plus précisément, parmi les personnes encore en études initiales, la part de celles qui ont obtenu un diplôme du supérieur croît avec l'âge et dépasse les 80 % à partir de 25 ans (Graphique 2).

**Graphique 2 - Part de personnes encore en études initiales et ayant obtenu un diplôme du supérieur**



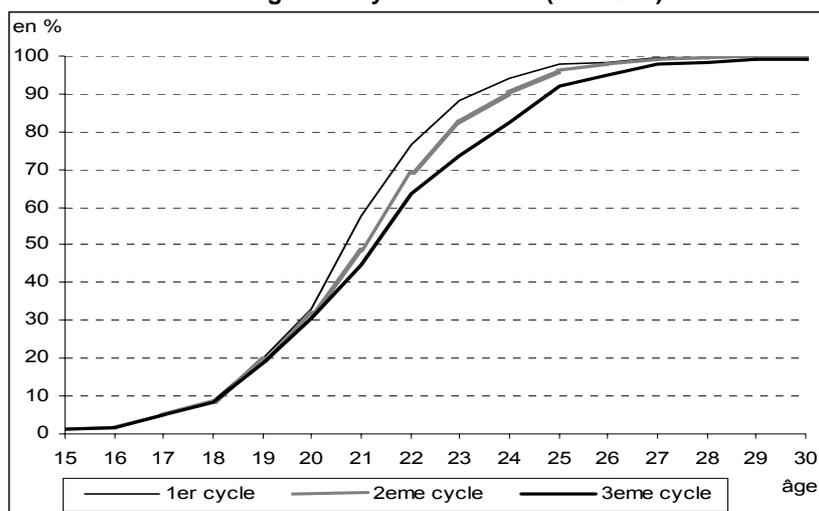
Champ : personnes en études initiales.  
Source : enquêtes Emploi 1990-2002, Insee.

La prise en compte de l'information sur le cursus des personnes en cours d'études permet de retenir un âge minimal plus élevé que si nous nous basions uniquement sur le fait d'être en études initiales ou non. Pour déterminer précisément l'âge minimal à retenir, regardons comment évolue la part de personnes appartenant au champ, c'est-à-dire celles pour lesquelles on a l'information sur le diplôme obtenu, selon l'âge. Pour chacun des cycles, le champ est défini de la façon suivante :

- champ 1<sup>er</sup> cycle : personnes ayant terminé leurs études ou celles en cours d'études et ayant un diplôme du supérieur.
- champ 2<sup>ème</sup> cycle : personnes appartenant au champ 1<sup>er</sup> cycle moins celles qui suivent une formation de 2<sup>ème</sup> cycle, mais n'ont pas de diplôme de ce niveau.
- champ 3<sup>ème</sup> cycle : personnes appartenant au champ 2<sup>ème</sup> cycle moins celles qui suivent une formation de 3<sup>ème</sup> cycle, mais n'ont pas de diplôme de ce niveau.

Il ressort du graphique 3 que la fraction des personnes hors champ devient minimale à partir de l'âge de 25 ans pour les 1<sup>er</sup> et 2<sup>ème</sup> cycle et 28 ans pour le 3<sup>ème</sup> cycle. C'est donc ces bornes inférieures que nous considérerons dans nos analyses.

**Graphique 3 - Part de personnes incluses dans le champ selon l'âge et le cycle considéré (en 2002<sup>18</sup>)**



Source : enquêtes Emploi 1990-2002, Insee.

### **Quelle borne supérieure en termes d'âge retenir ?**

La question de la délimitation du champ en termes d'âge se pose également pour les âges avancés, mais dans des termes différents. Il ne s'agit en effet pas d'un problème de champ (au-delà d'un certain âge, la quasi-totalité des personnes ont terminé leurs études initiales), mais de possibles biais de déclaration ou d'attrition sélective (mortalité différentielle selon le diplôme, mouvements migratoires). Baudelot, dans un article consacré à cette question<sup>19</sup>, constate en effet que plus les individus vieillissent, plus ils se déclarent diplômés. Il attribue cet effet aux « différentes appréciations que porteraient les individus sur le niveau de leurs propres qualifications (...) à mesure que (...) s'élèverait leur statut professionnel ». Un biais d'attrition, dû au fait qu'au fil des années une fraction d'une génération donnée n'est plus interrogée (en raison des décès, ou des migrations), joue dans le même sens que le biais précédent, à savoir un risque de déformation de la structure des diplômes vers des niveaux plus élevés quand on interroge une cohorte à un âge plus avancé<sup>20</sup>.

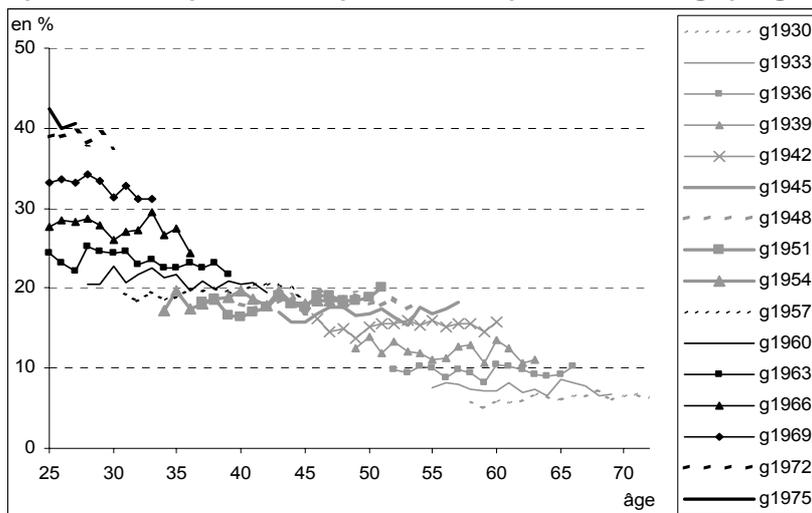
En compilant différentes enquêtes Emploi pour connaître la structure par diplôme d'une génération donnée, nous nous heurtons à ces possibles biais. En effet, nous faisons alors l'hypothèse que l'information sur le diplôme n'est pas affectée par l'âge auquel les individus d'une cohorte donnée sont interrogés. Il convient alors de vérifier si cette hypothèse est valide. Regardons pour cela la part de personnes diplômées du supérieur selon l'âge à l'enquête, pour différentes générations triennales.

<sup>18</sup> Etant donné l'élévation du niveau de diplôme, l'âge retenu en 2002 sera a fortiori valable pour les enquêtes antérieures.

<sup>19</sup> « L'âge rend-il plus savant ? Un exemple de biais de réponse dans les enquêtes », in *Les ménages, mélanges en l'honneur de Jacques Desabie*.

<sup>20</sup> En raison des risques de mortalité plus marqués des personnes peu diplômées (Monteil C., Robert-Bobée I. (2005), « Les différences sociales de mortalité : en augmentation chez les hommes, stables chez les femmes », Insee Première, n° 1025).

Graphique 4 - Part de personnes diplômées du supérieur selon l'âge par génération



Champ : personnes nées entre 1930 et 1977.  
 Source : enquêtes Emploi 1990-2002, Insee.

Il ressort du graphique 4 que les déclarations sur le diplôme semblent relativement stables quel que soit l'âge auquel elles sont recueillies<sup>21</sup>. Il ne s'avère donc pas nécessaire d'imposer une borne supérieure.

### **Champ retenu**

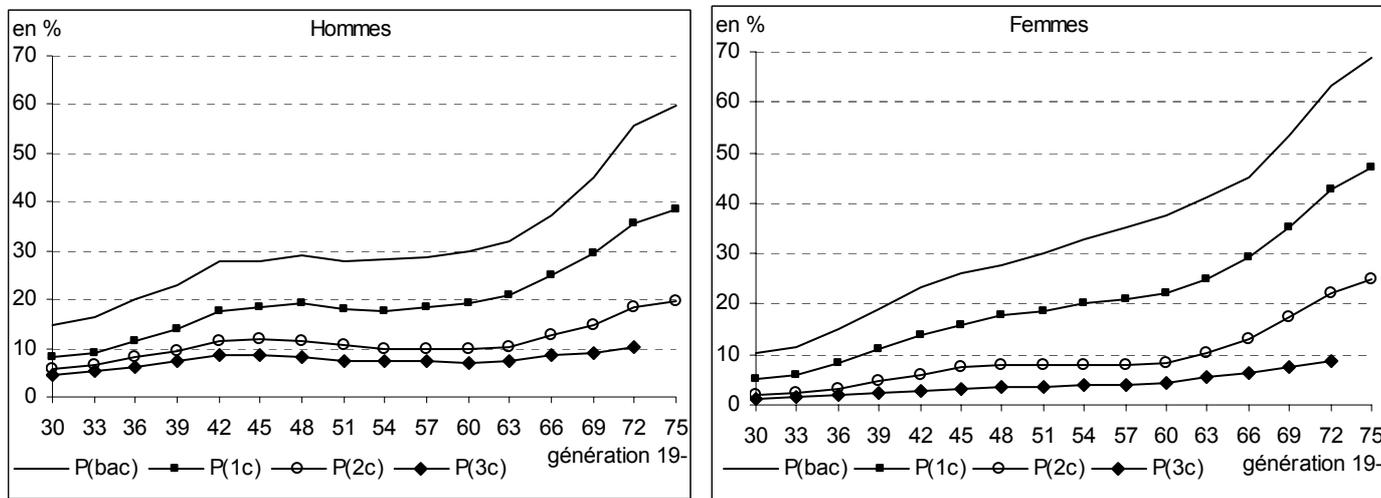
Au final, le champ retenu pour cette étude est donc constitué des personnes nées en France<sup>22</sup>, âgées d'au moins 25 ans pour l'étude des 1<sup>er</sup> et 2<sup>ème</sup> cycle et des filières du supérieur, ou d'au moins 28 ans pour le 3<sup>ème</sup> cycle, sans limite supérieure d'âge.

<sup>21</sup> Par construction, nous n'observons que des personnes âgées au plus de 72 ans, ce qui limite certainement les biais liés à la mortalité sélective.

<sup>22</sup> Afin de se restreindre autant que possible à des personnes ayant été scolarisées en France.

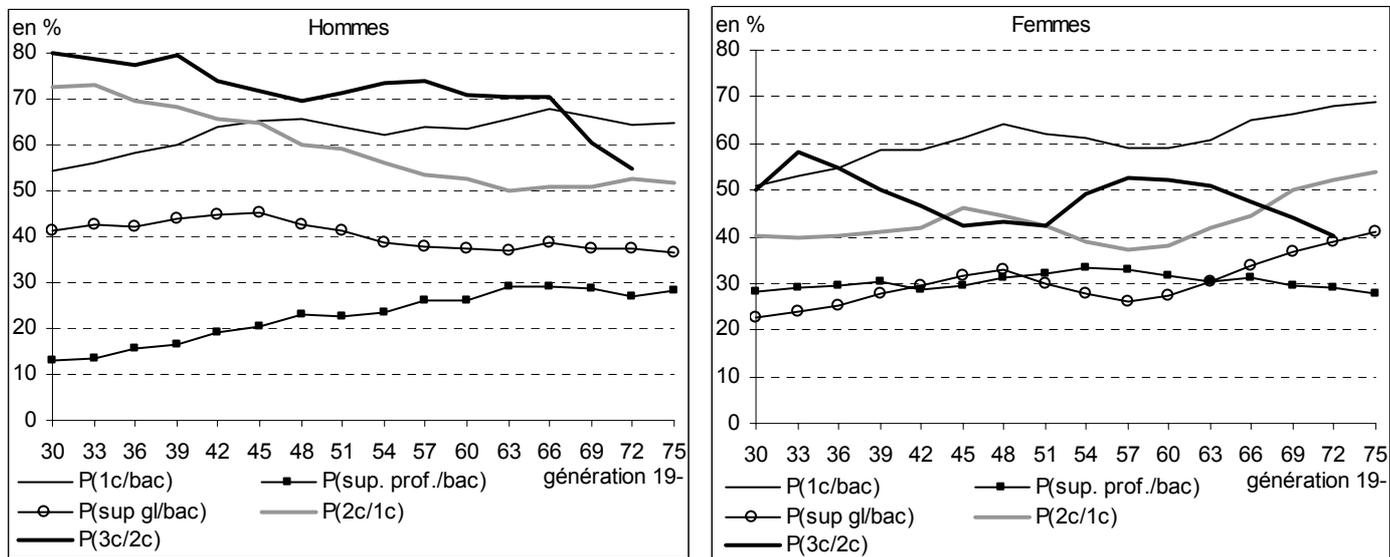
### Annexe 3 : Données de cadrage complémentaires

**A1 - Part de diplômés du supérieur selon les cycles d'enseignement**

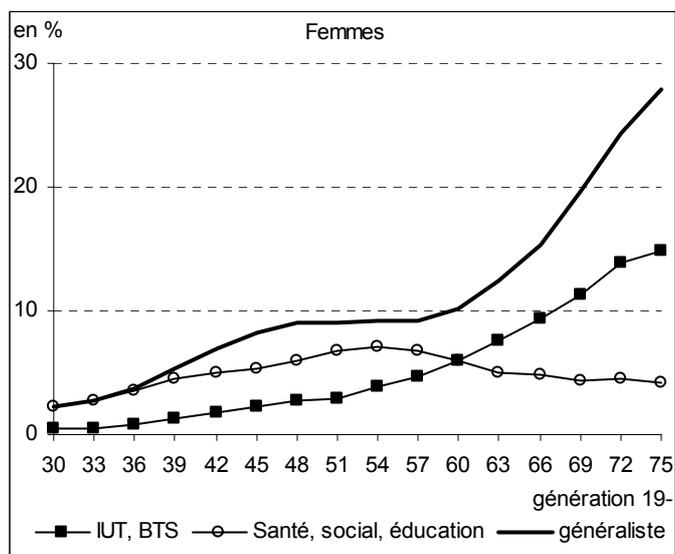
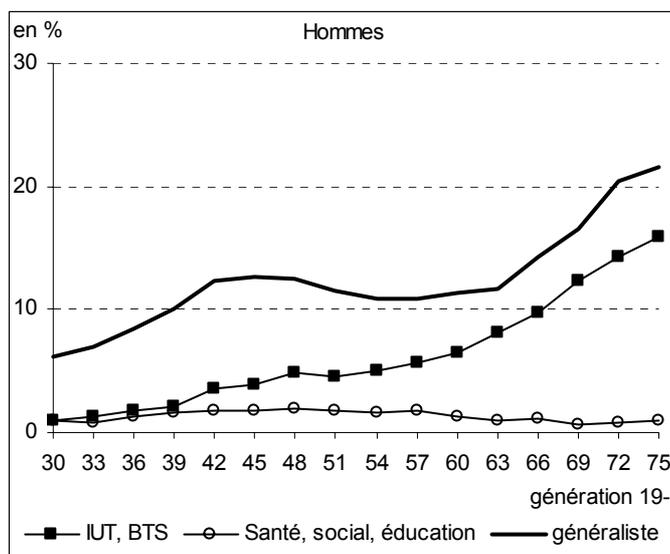


Note : P(1c) = probabilité d'être diplômé du supérieur, tous cycles confondus  
 P(2c) = probabilité d'être diplômé au moins du 2<sup>ème</sup> cycle du supérieur  
 P(3c) = probabilité d'être diplômé du 3<sup>ème</sup> cycle du supérieur

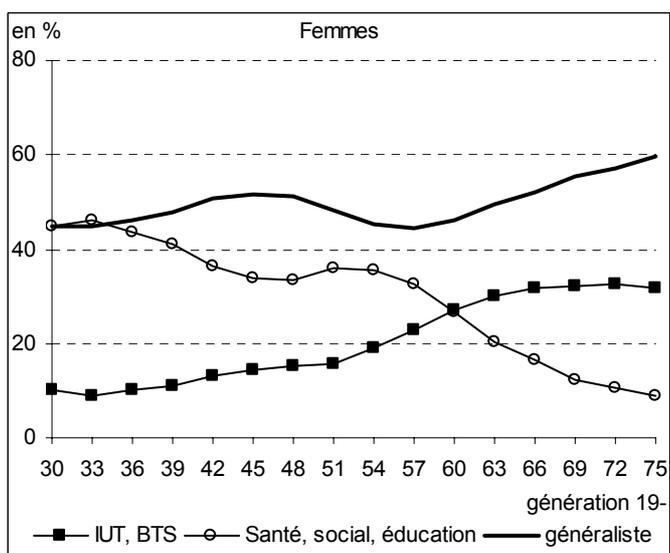
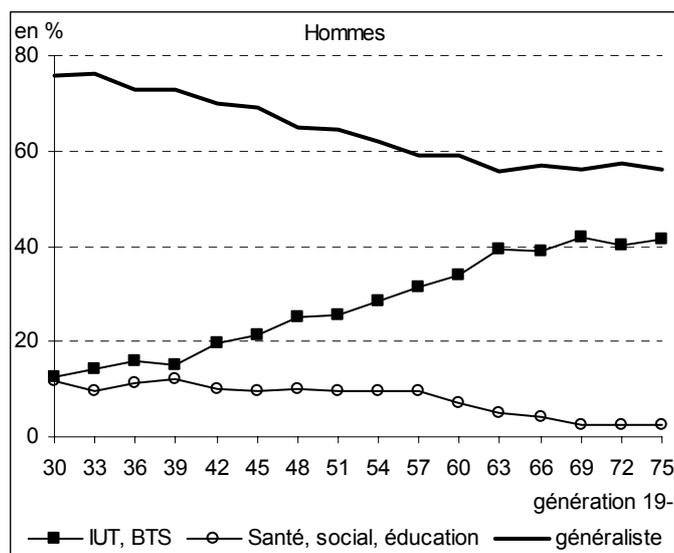
**A2 - Probabilités conditionnelles d'obtention de différents niveaux de diplômes**



### A3 - Part de diplômés du supérieur selon le type de diplôme

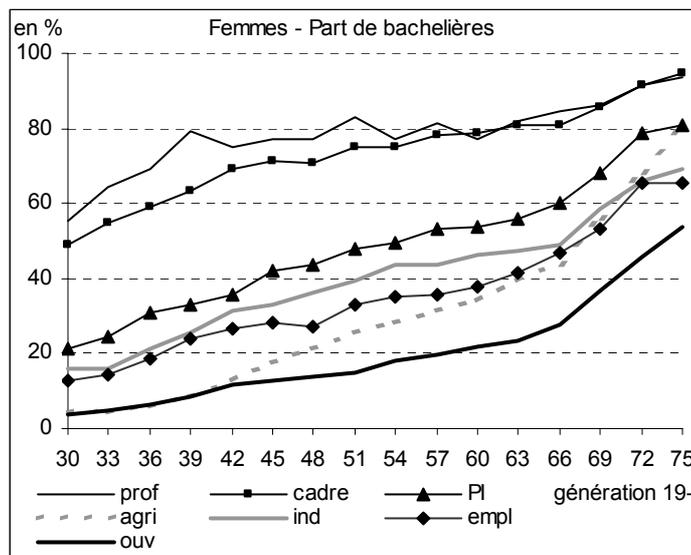
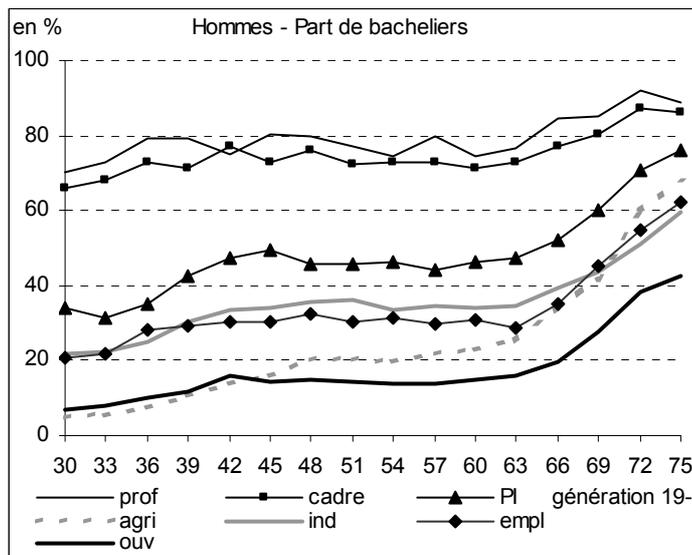


### A4 - Structure par type de diplôme des diplômés du supérieur



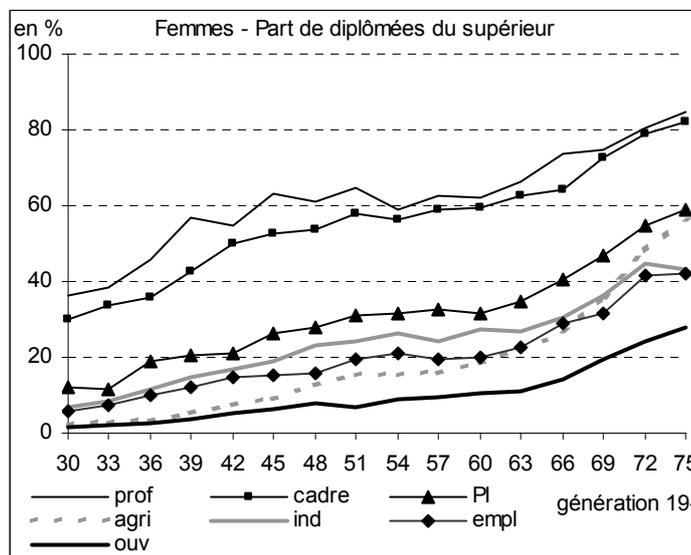
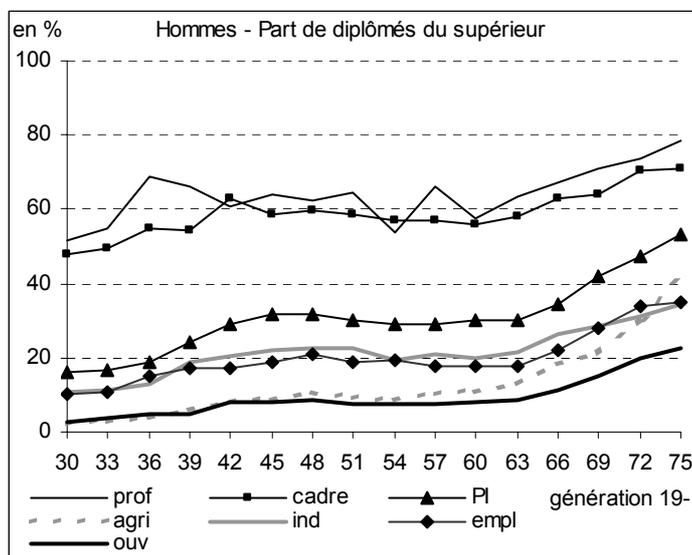
Champ : diplômés du supérieur.

**A5 - Part de bacheliers selon l'origine sociale**

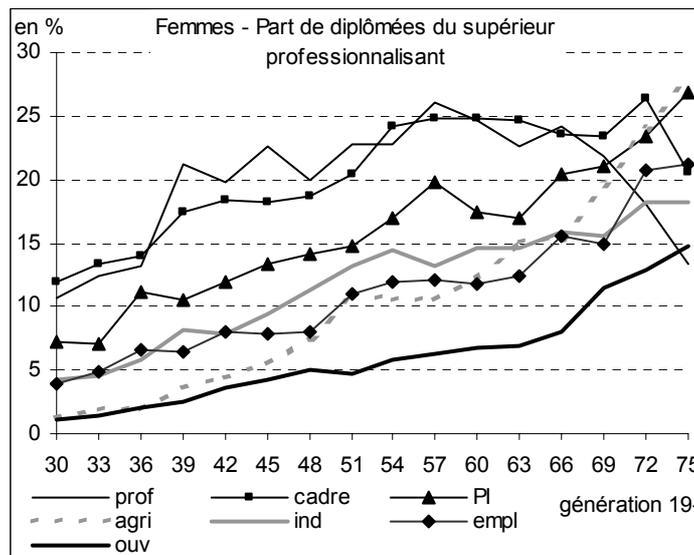
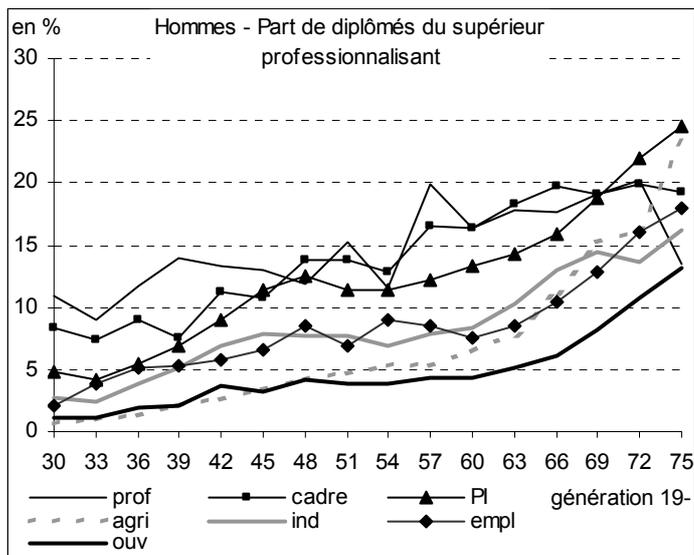


Note : prof = professeurs, instituteurs  
 cadre = cadres, professions libérales et chefs d'entreprise  
 PI = professions intermédiaires  
 agri= agriculteurs  
 ind = artisans et commerçants  
 empl = employés  
 ouv = ouvriers

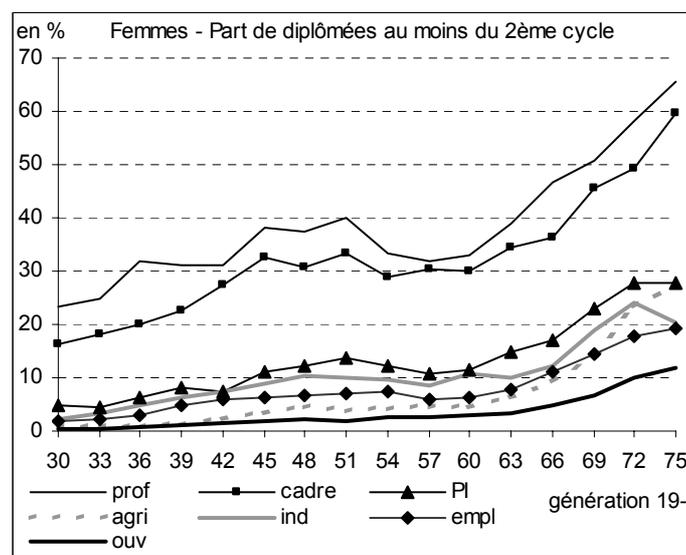
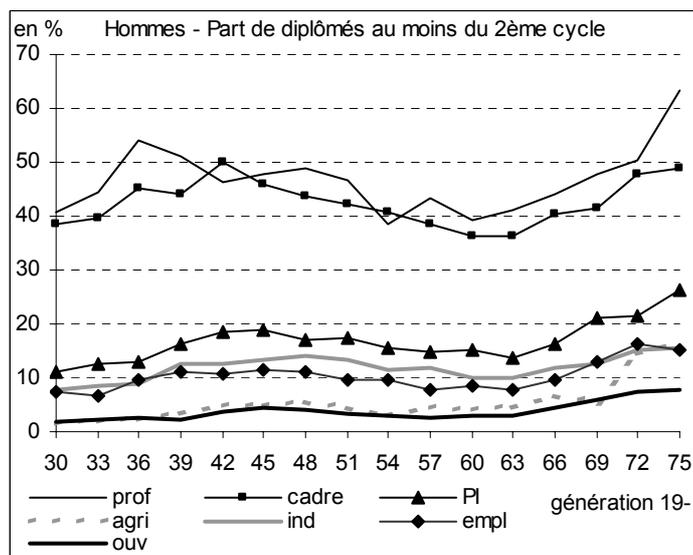
**A6 - Part de diplômés du supérieur selon l'origine sociale**



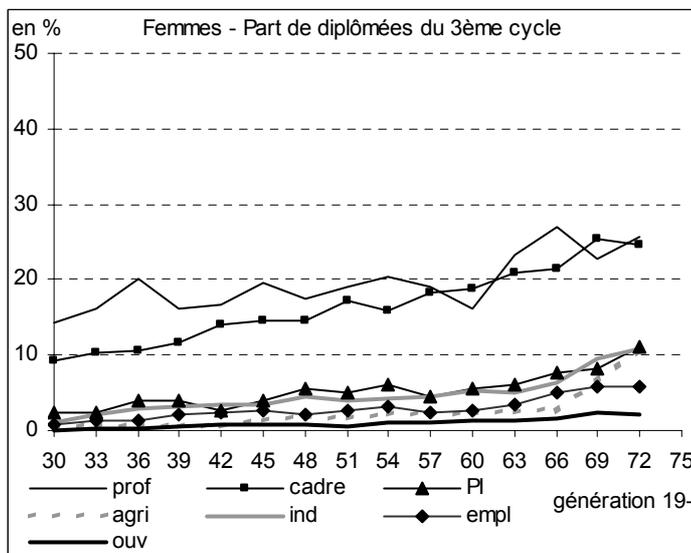
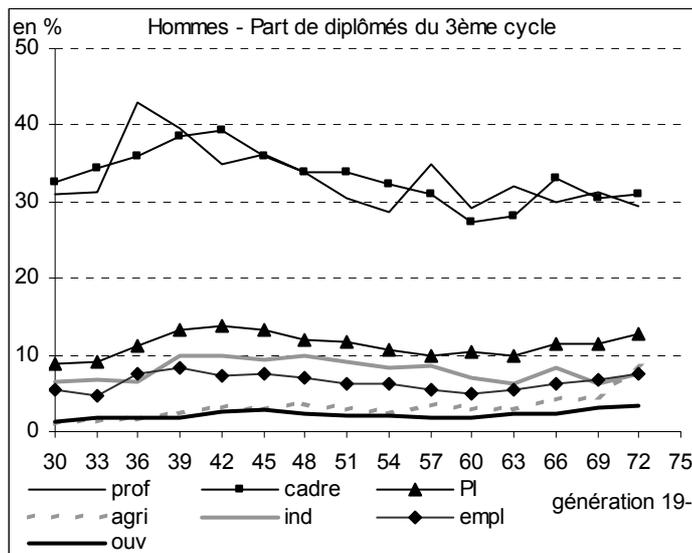
**A7 - Part de diplômés du supérieur professionnalisant selon l'origine sociale**



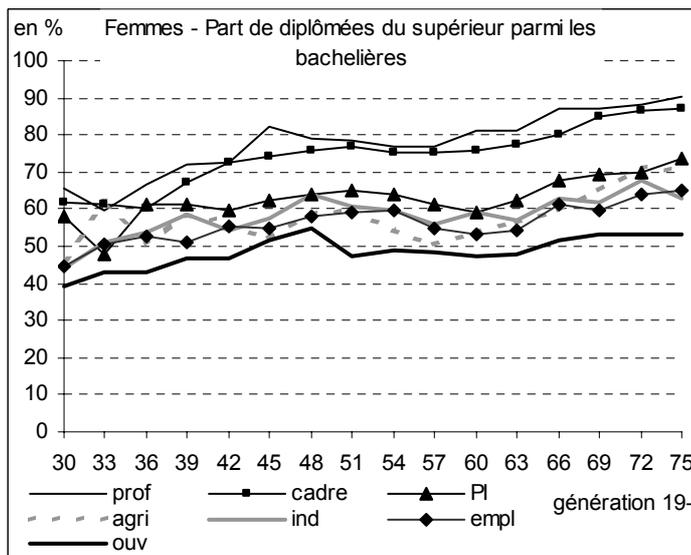
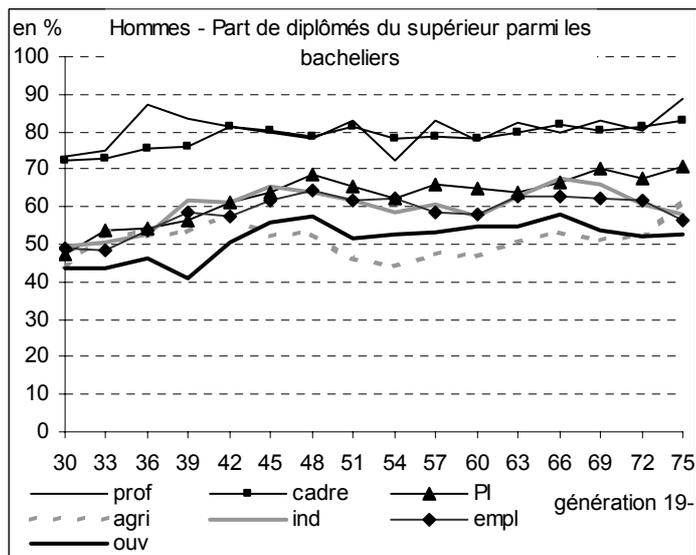
**A8 - Part de diplômés d'au moins un 2<sup>ème</sup> cycle du supérieur selon l'origine sociale**



**A9 - Part de diplômés d'un 3<sup>ème</sup> cycle du supérieur selon l'origine sociale**

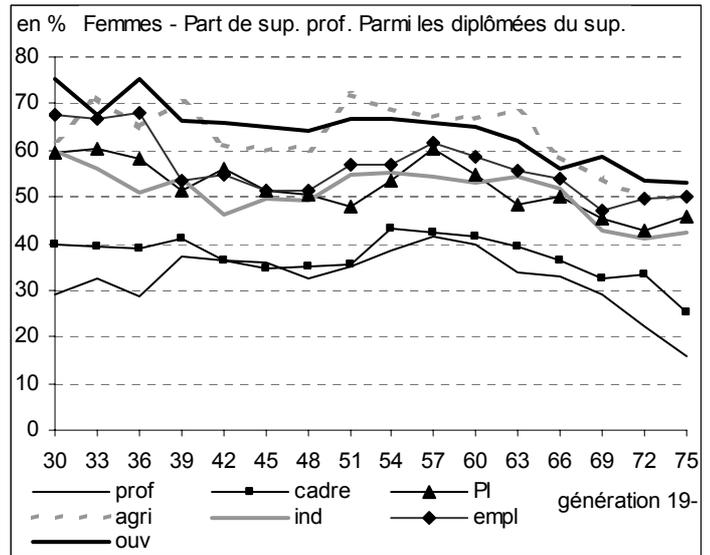
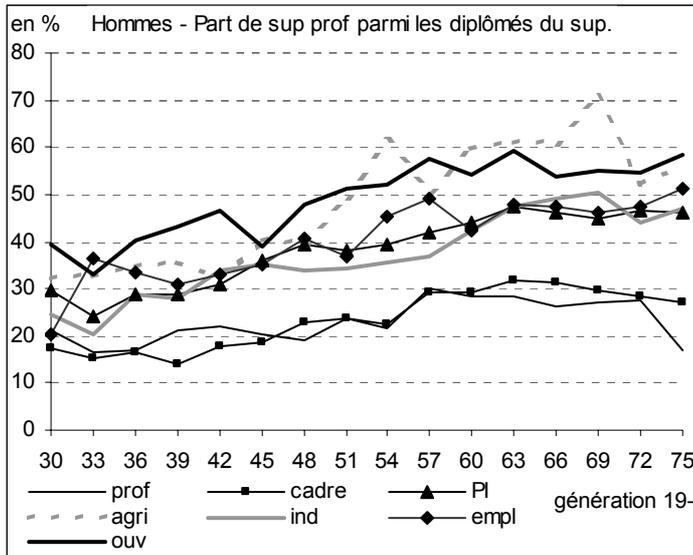


**A10 - Part de diplômés du supérieur parmi les bacheliers selon l'origine sociale**



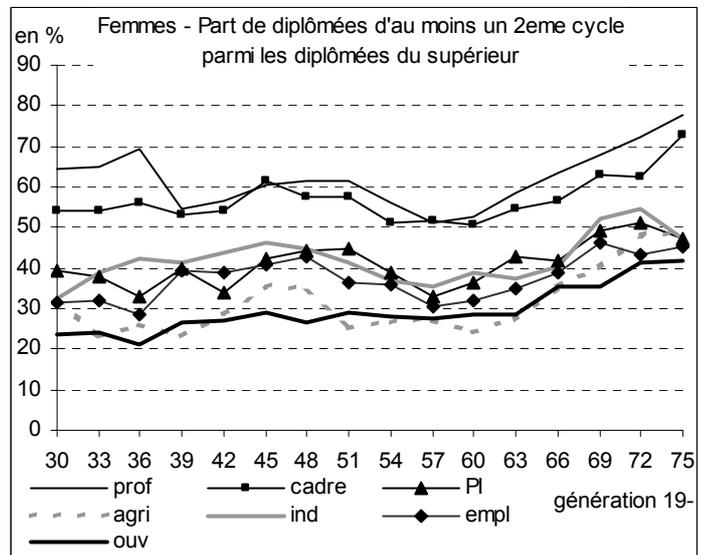
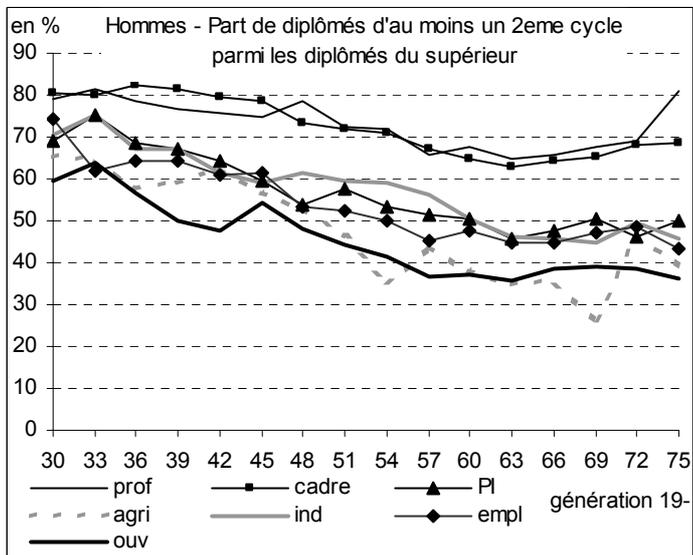
Champ : bacheliers

**A11 - Part de diplômés du supérieur professionnalisant parmi les diplômés du supérieur selon l'origine sociale**



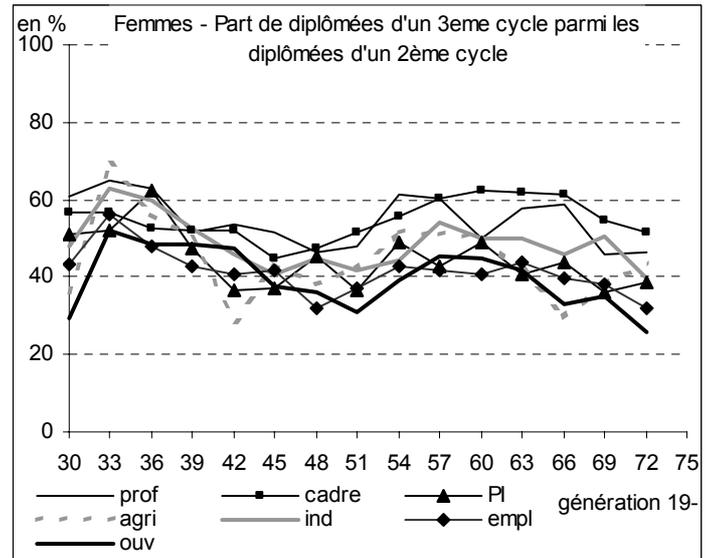
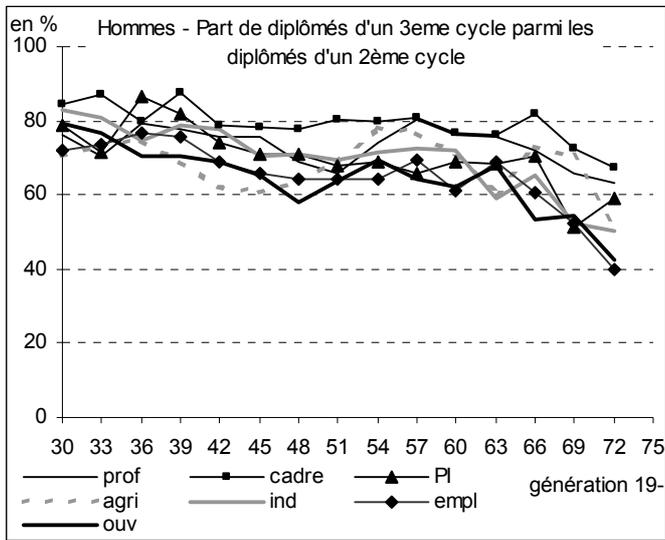
Champ : diplômés du supérieur.

**A12 - Part de diplômés d'au moins un 2ème cycle du supérieur parmi les diplômés du supérieur selon l'origine sociale**



Champ : diplômés du supérieur

**A13 - Part de diplômés du 3<sup>ème</sup> cycle parmi les diplômés d'au moins un 2<sup>ème</sup> cycle du supérieur selon l'origine sociale**



Champ : personnes ayant un diplôme supérieur ou égal au 2<sup>ème</sup> cycle

## Annexe 4 : Méthodologie

### *Les inégalités scolaires sont mesurées avec la métrique logistique*

La métrique que nous utilisons pour comparer les probabilités entre elles est l'**échelle logistique**. Cette mesure a la caractéristique d'adapter la mesure de l'évolution d'une probabilité au niveau initial de cette probabilité. Pour cette raison, elle permet de comparer l'évolution de probabilités variant sur un large éventail. Techniquement, adopter cette échelle logistique revient à raisonner linéairement sur la variable latente définie par  $y = \ln\left(\frac{p}{1-p}\right)$ . On appelle parfois cette variable latente « propension ».

Mesurer l'évolution des probabilités  $p$  revient à calculer les écarts en différence de cette variable latente. Ainsi, la mesure de l'évolution d'une probabilité  $p$  entre deux périodes  $t_1$  et  $t_2$  se fera par :

$$y_{t_2} - y_{t_1} = \ln\left(\frac{p_{t_2}}{1-p_{t_2}}\right) - \ln\left(\frac{p_{t_1}}{1-p_{t_1}}\right).$$

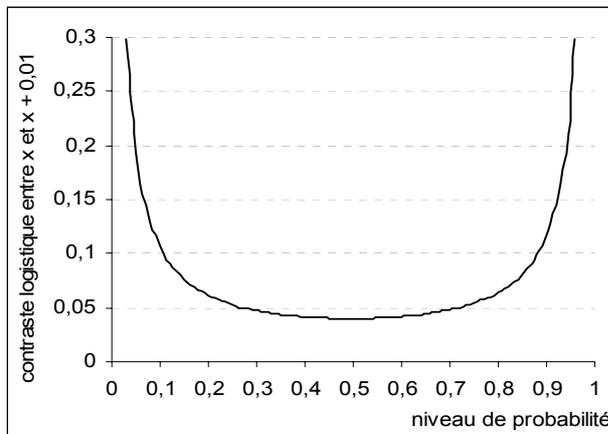
De même, la mesure du contraste des probabilités observées au sein de deux groupes  $a$  et  $b$  se fait par le contraste logistique :

$$y_a - y_b = \ln\left(\frac{p_a}{1-p_a}\right) - \ln\left(\frac{p_b}{1-p_b}\right).$$

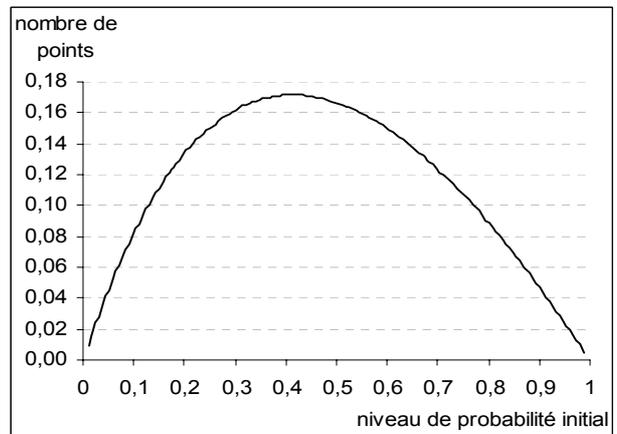
Plus précisément, le graphique A montre la valeur du contraste logistique correspondant à l'évolution de 1 point de probabilité selon le niveau initial de cette probabilité. On peut y lire qu'un accroissement de 1 point se traduit par un contraste logistique beaucoup plus élevé pour les probabilités proches de 0 ou de 1. Pour les niveaux de probabilité intermédiaires, situés entre 25 et 75 %, l'évolution de 1 point correspond à un contraste à peu près constant. Cette mesure traduit donc les deux idées intuitives suivantes : celle que le passage de 3 à 4 % ou de 90 à 91 % représente une évolution plus forte que le passage de 20 à 21 % ou de 60 à 61 % ; celle aussi que l'évolution de 20 à 21 % et celle de 60 à 61 % représentent la même évolution.

Une manière alternative est de tracer la courbe « iso-contraste » : elle représente de combien de points doit augmenter chaque probabilité pour que l'évolution mesurée par l'échelle logistique soit constant. Prenons l'exemple d'un odds ratio de 2 (et donc d'un contraste de 0,69), on peut lire sur le graphique B qu'un accroissement de 5 points d'une probabilité initialement à 5 % (soit son passage de 5 à 10 %) représente autant qu'un accroissement de 8 points d'une probabilité à 10 % ou encore qu'un accroissement de 17 points d'une probabilité à 40 %.

**Graphique A : Valeur de l'évolution de 1 point de probabilité mesurée par le contraste logistique**



**Graphique B : courbe « iso-contraste » : accroissement nécessaire (en points) pour que le contraste logistique (mesurant l'évolution de la probabilité) soit égal à 2**



### **Métrique logistique et odds ratio**

Quand il ne s'agit que de déterminer si le contraste entre deux populations s'est réduit ou non dans le temps, on préfère généralement utiliser l'exponentielle du contraste logistique soit :

$$\exp(y_a - y_b) = \frac{p_a(1-p_b)}{(1-p_a)p_b} \text{ que l'on appelle odds ratio ou rapport de chances relatives.}$$

La fonction exponentielle étant monotone croissante, une atténuation de ce odds ratio est bien équivalente à une atténuation du contraste logistique.

Dans le cas général, l'interprétation la plus simple du odds ratio est la suivante : soient  $p_1$  et  $p_2$  les probabilités d'avoir une caractéristique donnée (par exemple d'être diplômé en sciences humaines) dans deux sous-populations  $a$  et  $b$ , prenons deux personnes au hasard, la première appartenant à la sous-population  $a$  et la deuxième à la sous-population  $b$ . Un odds ratio valant  $k$  signifie qu'il y a  $k$  fois plus de chances que ce soit la personne du premier groupe qui ait cette caractéristique et l'autre non, plutôt que l'inverse.

Lorsque, comme c'est le cas dans ce travail, on raisonne sur la probabilité d'être au moins diplômé d'un niveau  $n$ , l'interprétation du odds ratio se simplifie. En effet, fixons un niveau de diplôme et scindons la population en deux : les personnes qui ont au moins ce niveau de diplôme, et les autres. Le odds ratio, ou le contraste logistique, a alors l'intéressante propriété d'être une mesure du classement scolaire relatif de deux individus. C'est le rapport entre les chances que la personne appartenant au premier groupe social soit plus diplômée que celle issue du deuxième groupe et les chances qu'on soit dans la situation inverse (que ce soit la personne du deuxième groupe qui soit la plus diplômée).

### **La modélisation des tableaux de contingence pour résumer les évolutions**

L'étude de l'évolution des inégalités sociales à l'école peut tout à fait se faire en regardant l'évolution temporelle des contrastes logistiques (ou des odds ratio) entre groupes sociaux. Si l'on scinde la population en sept milieux sociaux, une difficulté supplémentaire émerge du fait que l'on étudie l'évolution temporelle de 21 contrastes

logistiques. Il devient dès lors difficile d'avoir une vision synthétique de l'évolution globale des inégalités sociales. On utilise alors des modèles dans le but de résumer ces évolutions.

### ***Des modèles log linéaires qui font d'emblée l'hypothèse que les inégalités n'ont pas évolué***

Les modèles log-linéaires, les premiers à avoir été développés, sont utiles pour tester l'indépendance conditionnelle entre variables. Soit par exemple  $m_{ijg}$  le nombre de personnes du groupe social  $i$  et de la génération  $g$  ayant le niveau de diplôme  $j$ . Le modèle de la forme :

$$\log m_{ijg} = \lambda + \lambda_i + \lambda_j + \lambda_g + \lambda_{ig} + \lambda_{jg} + \lambda_{ij}$$

avec les contraintes :

$$\left\{ \begin{array}{l} \sum_i \lambda_i = \sum_j \lambda_j = \sum_g \lambda_g = 0 \\ \sum_i \lambda_{ij} = \sum_j \lambda_{ij} = \sum_i \lambda_{ig} = \sum_g \lambda_{ig} = \sum_j \lambda_{jg} = \sum_g \lambda_{jg} = 0 \end{array} \right.$$

spécifie qu'il existe une relation entre le fait d'être diplômé et l'appartenance à un groupe social (par le paramètre d'ordre 2  $\lambda_{ij}$ ), mais que ce lien statistique est constant d'une génération à l'autre. En effet, le seul lien entre les variables de diplôme ( $j$ ) et d'origine sociale ( $i$ ) est le paramètre  $\lambda_{ij}$ , constant pour toutes les générations. Dans ce modèle, les paramètres  $\lambda_i$  et  $\lambda_{ig}$  s'ajustent à la taille des groupes sociaux, les paramètres  $\lambda_j$  et  $\lambda_{jg}$  au nombre des diplômés, et le paramètre  $\lambda_g$  à la taille des générations.

Le modèle est postulé vrai en population générale, mais est ajusté à partir d'effectifs d'enquête. On suppose que ces effectifs sont issus d'un tirage aléatoire en population générale avec une taille d'échantillon fixée a priori à  $N$ . Les effectifs observés  $n_{ijg}$  suivent alors une loi multinomiale de densité :

$$f(n_{ijg}/N; \pi_{ijg}) = \left( \frac{N!}{\prod_{ijg} n_{ijg}} \right) \prod_{ijg} (\pi_{ijg})^{n_{ijg}} \text{ avec } \pi_{ijg} = \frac{m_{ijg}}{N}$$

Le modèle est ajusté par maximum de vraisemblance par le logiciel LEM développé par J. K. Vermunt (1997).

On peut vérifier que le modèle spécifié ci-dessus revient à écrire que la probabilité d'avoir le diplôme  $j$  pour un enfant du milieu social  $i$  de la génération  $g$  ( $\pi_{j/ig}$ ) s'écrit :

$$\pi_{j/ig} = \frac{1}{1 + \exp(-2\lambda_j - 2\lambda_{ij} - 2\lambda_{jg})}$$

ce qui se réécrit

$$\log \left( \frac{\pi_{j/ig}}{1 - \pi_{j/ig}} \right) = 2(\lambda_j + \lambda_{ij} + \lambda_{jg})$$

et conduit à

$$\Delta y(i, i' / j, g) = \log \left( \frac{\pi_{j/i_g}}{1 - \pi_{j/i_g}} \right) - \log \left( \frac{\pi_{j/i'_g}}{1 - \pi_{j/i'_g}} \right) = 2 (\lambda_{ij} - \lambda_{i'j})$$

(où  $\Delta y(i, i' / j, g)$  est le contraste logistique entre les milieux  $i$  et  $i'$ ).

Ce modèle est donc équivalent à un modèle logistique avec les variables génération et origine sociale comme explicatives, non croisées. Il repose sur l'hypothèse que les contrastes logistiques n'ont pas évolué dans le temps et présente l'inconvénient de ne pas permettre de tester cette hypothèse.

### ***L'apport du modèle log-multiplicatif de Xie : ajuster un modèle d'évolution homothétique***

Les modèles log-multiplicatifs sont pour cela plus puissants. Ils permettent de tester si le lien entre origine sociale et niveau scolaire se déforme dans le temps. Parmi eux figure le modèle d'évolution uniforme des chances relatives, proposé par Xie (1992) et introduit en France par Vallet (1999) qui s'écrit :

$$\log m_{ijg} = \lambda + \lambda_i + \lambda_j + \lambda_g + \lambda_{ig} + \lambda_{jg} + \beta_g \psi_{ij}$$

sous les contraintes :

$$\begin{cases} \sum_i \lambda_i = \sum_j \lambda_j = \sum_g \lambda_g = 0 \\ \sum_i \psi_{ij} = \sum_j \psi_{ij} = \sum_i \lambda_{ig} = \sum_g \lambda_{ig} = \sum_j \lambda_{jg} = \sum_g \lambda_{jg} = 0 \\ \beta_{g_0} = 1 \end{cases}$$

Ce modèle suppose qu'il existe une structure de base des inégalités scolaires, qui s'exprime avec un facteur d'intensité ( $\beta_g$ ) à chaque génération. Entre deux générations successives, l'ensemble des contrastes scolaires sont amplifiés, ou atténués, selon le même facteur d'intensité ( $\beta_g$ ).

En effet, on montre que :

$$\log \left( \frac{\pi_{j/i_g}}{1 - \pi_{j/i_g}} \right) = 2 (\lambda_j + \lambda_{jg} + \beta_g \psi_{ij})$$

$$\Delta y(i, i' / j, g) = \beta_g * 2 (\psi_{ij} - \psi_{i'j})$$

Le contraste social des taux d'obtention du diplôme  $j$  pour les milieux  $i$  et  $i'$  au sein de la génération  $g$  se décompose dans ce modèle entre un facteur indépendant de la génération  $-2^*(\psi_{ij} - \psi_{i'j})$  et un facteur d'intensité lié à la génération ( $\beta_g$ ), mais identique pour tous les milieux sociaux pris deux à deux. Il doit donc varier dans le même sens d'une génération à l'autre quelles que soient les origines sociales entre lesquelles il est calculé. Ce modèle sera adapté aux données de destinée scolaire en fonction de l'origine sociale s'il y a une tendance générale d'une génération à l'autre, soit dans le sens d'un renforcement des inégalités, soit dans le sens d'une démocratisation qualitative (c'est-à-dire tous les *odds ratios* évoluent dans le même sens, et non dans un sens qui dépend des origines sociales considérées).

### **Les statistiques d'adéquation des modèles**

Plusieurs statistiques permettent de juger de l'adéquation des modèles aux données. La statistique du rapport des vraisemblances (notée  $L^2$ ) permet tout d'abord, en regard au nombre de degrés de liberté du modèle ( $ddl$ ), de produire un test d'acceptation du modèle. Elle s'écrit :

$$L^2 = 2 \sum_i \sum_j \sum_g n_{ijg} \log\left(\frac{n_{ijg}}{m_{ijg}}\right)$$

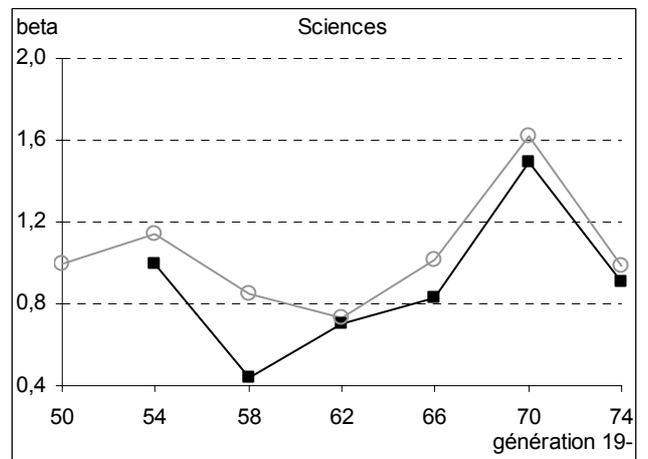
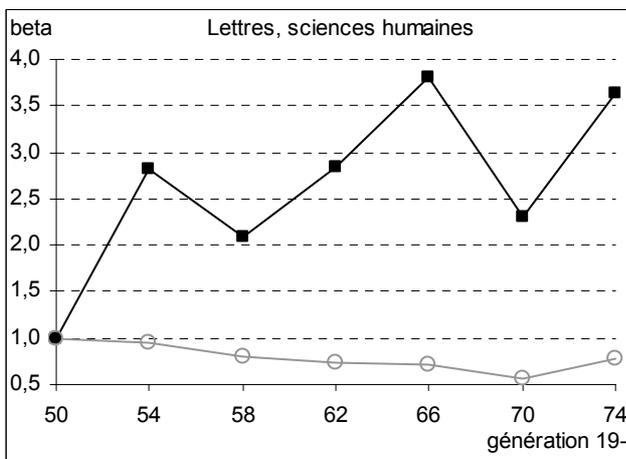
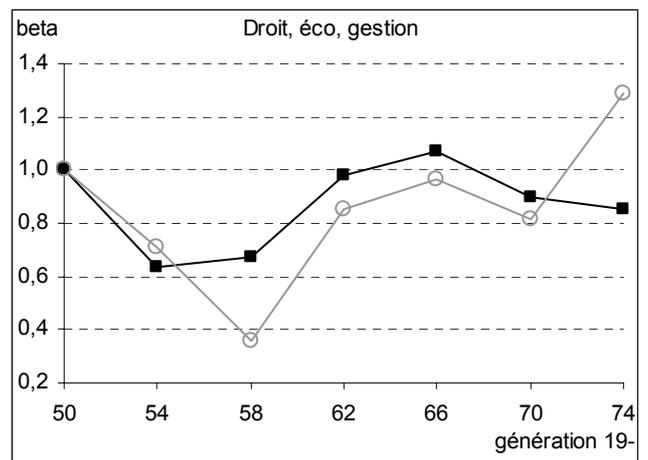
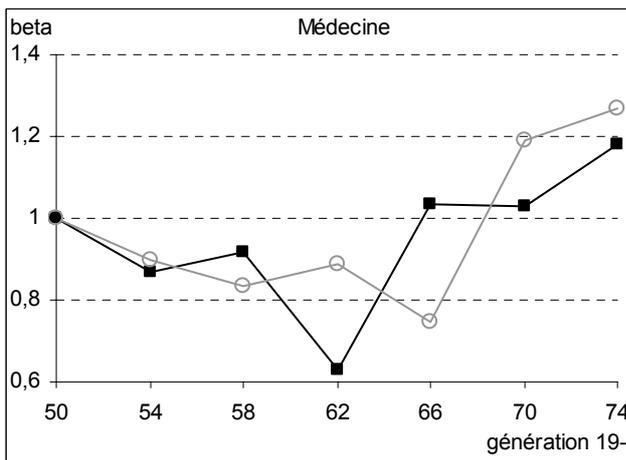
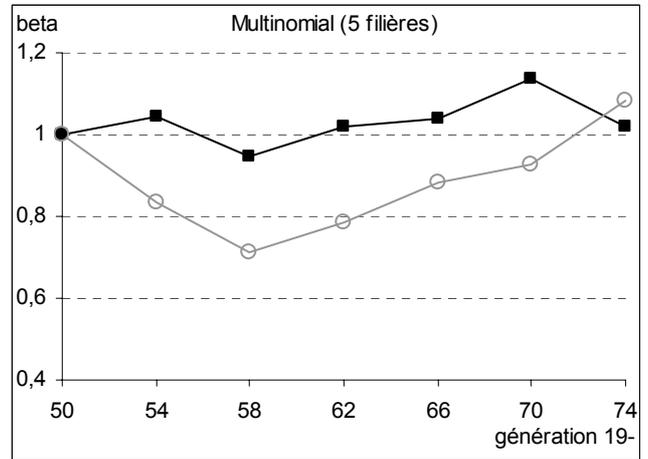
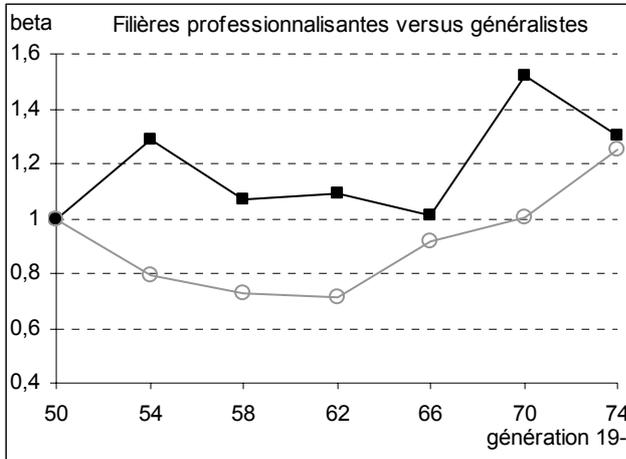
où  $n_{ijg}$  et  $m_{ijg}$  sont respectivement les effectifs observés et estimés.  $L^2$  suit une loi du  $\chi^2$  avec un degré de liberté égal au nombre de degrés de liberté du modèle. Toutefois, il faut garder à l'esprit que ce genre de test est sensible à la taille de l'échantillon sur lequel on travaille (plus l'effectif est grand, plus le test va être exigeant).

On peut lui préférer des indicateurs qui prennent en compte la taille de l'échantillon. Mais alors, il faut plutôt se placer dans une optique de comparaison entre modèles. Le premier indicateur est le **Bayesian Information Criterion (BIC)** qui vaut  $BIC=L^2-dl/\log(N)$ , où  $N$  est la taille de l'échantillon. Mais ce dernier ne permet pas d'accepter ou de refuser un modèle pris isolément. Le second indicateur de la qualité du modèle est **l'indice de dissimilarité (D)**. Il indique le pourcentage de personnes mal classées par le modèle. Là encore, il permet difficilement d'avoir une opinion absolue sur la pertinence du modèle au regard des données.

### **Test de significativité des différences des $\beta_g$ du modèle d'évolution uniforme des chances relatives**

On peut également établir des tests de significativité des tendances dégagées par le  $\beta_g$ . On estime ce modèle en imposant l'égalité des coefficients  $\beta_g$  de deux générations successives. La double différence entre les log-vraisemblances des modèles contraint et non contraint, comparée au fractile d'ordre 95 % d'une loi du  $\chi^2$  à un degré de liberté permet d'établir la significativité de la différence entre les deux paramètres du modèle non contraint. On peut généraliser cette démarche pour tester la significativité de l'égalité entre  $n$   $\beta_g$  différents ; il faut alors comparer la statistique au fractile d'un  $\chi^2$  à  $n-1$  degrés de liberté

### Annexe 5 : Paramètres d'intensité des modèles d'évolution uniforme dans l'analyse par filière



Source : Enquêtes Emploi 1996 à 2002.

Note : Les courbes en noir avec des carrés concernent les garçons ; celles en gris avec des cercles les filles. Un paramètre qui décroît signifie que la polarisation de la filière s'atténue.

Champ : personnes diplômées du supérieur.

## Annexe 6 : Mise en œuvre pratique du test de significativité des coefficients $\beta_g$ entre deux modèles différents

### Principe

Si l'égalité de deux coefficients  $\beta_g$  d'un même modèle peut se tester aisément en recourant à un test de vraisemblance, celle entre deux coefficients  $\beta_g$  issus de deux modèles différents (portant par exemple sur des sous-populations différentes) nécessite d'estimer les écarts-type des coefficients. L'estimation par bootstrap des écarts-type de  $\beta_g$  est en pratique assez lourde à mettre en œuvre car elle nécessite l'ajustement de T+1 tableaux de contingence (c'est-à-dire des tableaux d'effectifs) origine sociale  $\times$  obtention d'un diplôme de cycle  $n$ , où T est le nombre de tirages (ici T=100), auquel on ajoute l'ajustement sur le tableau d'origine, directement issu des données d'enquête. Une utilisation standard de LEM, logiciel utilisé ici pour estimer nos différents modèles, oblige à exécuter le programme autant de fois qu'on a de tableaux de contingence, soit T+1 fois dans le cas présent. L'idée est alors de construire une matrice constituée par nos T+1 tableaux telle que son estimation soit équivalente à l'estimation successive des T+1 tableaux pris séparément. Pour ce faire, nous utilisons le fait qu'une case du tableau de contingence sans effectif ne contribue pas à la vraisemblance. Une mise en forme particulière de nos données, sous la forme d'une matrice par blocs où le nombre de périodes et le nombre de catégories sociales sont dupliqués autant de fois que l'on a de tableaux et où l'on place de façon judicieuse des zéros, permet d'obtenir en une seule exécution les  $\beta_g$  issus de nos T+1 modèles différents<sup>23</sup>

Ajuster le modèle suivant :

$$\log m_{ijg} = \lambda + \lambda_i + \lambda_j + \lambda_g + \lambda_{ig} + \lambda_{jg} + \beta_g \psi_{ij}, \text{ avec } m_{ijg} \text{ effectif observé}$$

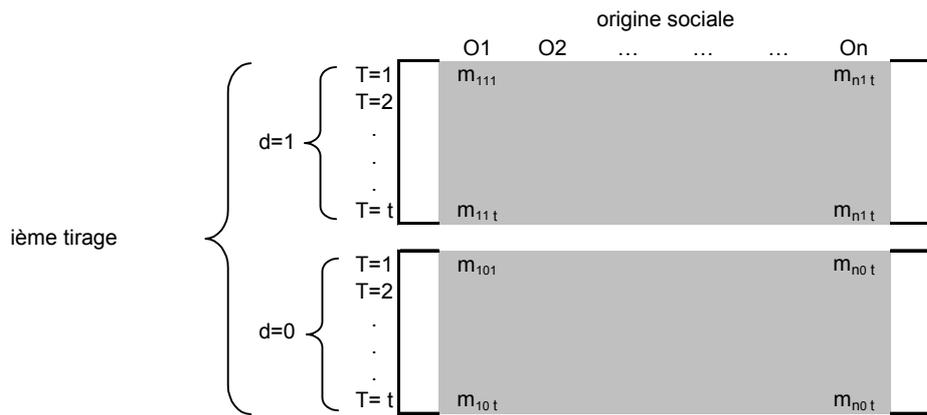
sous les contraintes :

$$\left\{ \begin{array}{l} \sum_i \lambda_i = \sum_j \lambda_j = \sum_g \lambda_g = 0 \\ \sum_i \psi_{ij} = \sum_j \psi_{ij} = \sum_i \lambda_{ig} = \sum_g \lambda_{ig} = \sum_j \lambda_{jg} = \sum_g \lambda_{jg} = 0 \\ \beta_{g_1} = 1 \end{array} \right.$$

successivement sur les données de deux tirages, chacune présentée de la façon suivante :

---

<sup>23</sup> En raison de la taille de la matrice obtenue, les données doivent être enregistrées dans un fichier texte (extension .fre) appelé depuis LEM.



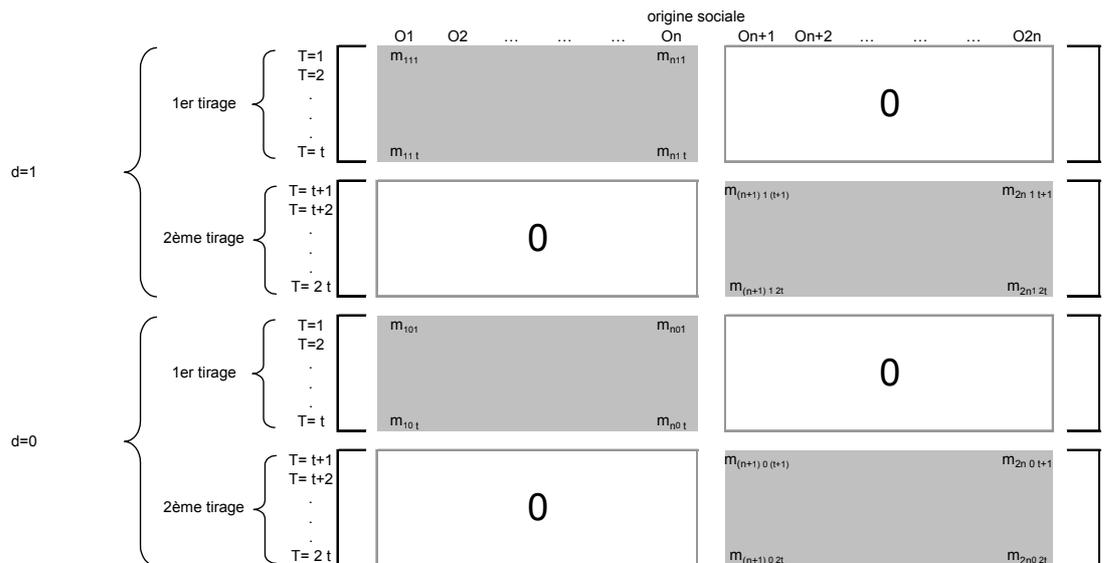
Est alors équivalent à estimer le sur-modèle

$$\log m_{ijg} = \lambda + \lambda_i + \lambda_j + \lambda_g + \lambda_{ig} + \lambda_{jg} + \beta_g \psi_{ij}$$

sous les contraintes :

$$\begin{cases} \sum_i \lambda_i = \sum_j \lambda_j = \sum_g \lambda_g = 0 \\ \sum_i \psi_{ij} = \sum_j \psi_{ij} = \sum_i \lambda_{ig} = \sum_g \lambda_{ig} = \sum_j \lambda_{jg} = \sum_g \lambda_{jg} = 0 \\ \beta_{g_1} = 1, \beta_{g_{t+1}} = 1 \end{cases}$$

sur les données présentées de la façon suivante :



Le programme suivant effectue les k tirages du bootstrap, construit les matrices par blocs, estime le sur modèle et calcule les écarts-type des coefficients.

### Principales étapes et programme

Les principales étapes concrètes sont :

#### 1) Produire une matrice par blocs contenant les T+1 tableaux de contingence origine sociale × obtention d'un diplôme de cycle $n$

En complétant un programme de tirage avec remise (Afsa, 2003), on construit différentes macro SAS permettant de mettre en forme les données.

**%DATABOOT** : macro produisant la matrice par blocs contenant les T+1 tableaux de contingence origine sociale × obtention d'un diplôme de cycle  $n$  sur données observées

```

/*****
*****/
/*Macro sortant les effectifs observés avec bootstrap      */
/*cas sur données observées                               */
/*****
*****/
/*gene = variable définissant les générations            */
/*naia_inf = année définissant l'année de naiss         */
/* minimale de notre champ                               */
/*naia_sup = année définissant l'année de naiss         */
/* maximale de notre champ                               */
/*age = age inf du champ (25 ou 28 pour le 3eme cycle)  /
/*sexe = 1 ou 2                                          */
/*champ = variable définissant le champ d'interet      */
/*dip      =      variable      définissant      le      diplôme
d'interet,dichotomique/*t = nombre de tirages, si t=T, on a au
final T+1 groupes de données * car T=0 = données observées sur
l'échantillon initial*/
/*****
*****/
/*La macro crée un fichier txt dans D\HAS4B7 de nom
KRONbtirage+1
/* Ex : kro 101                                          */
/*****
*****/
/*qd on travaille avec gene4_1977 '1950'<=naia<='1977' */
/*qd on travaille avec gene3_1977 naia débute en 1948   */
/*****
*****/

```

**%macro** databoot (gene=,naia\_inf=,naia\_sup=,age=,sexe=,champ=,dip=,t=);

```

data ent;
set eeis.eea1990_2002 (where=("&naia_inf."<=naia<="&naia_sup." & cspere ne ''
& &champ.=1 & s="&sexe." & &age.<=age));
keep nopers &dip. &gene. cspere n;
run;

proc summary data=ent nway;
var n;
class &dip. &gene. cspere;
output out=titi sum=;
run;

proc sort data=titi ; by &dip. &gene.; run;

```

```

proc transpose data=titi out=eff (drop = _NAME_) prefix =
cspere;
by &dip. &gene.;
var n;
id cspere;
run;

data eff;
T=0;
set eff;
if cspere1=. then cspere1=0;
if cspere2=. then cspere2=0;
if cspere3=. then cspere3=0;
if cspere4=. then cspere4=0;
if cspere5=. then cspere5=0;
if cspere6=. then cspere6=0;
if cspere7=. then cspere7=0;
run;

%do k=1 %to &t.;
  %put =====> &k eme iteration;
  data tab_boot;
  retain ii 0;
  num_obs=int(ranuni(12345+&k.)*tot_obs)+1;
  set ent (keep=nopers) point=num_obs nobs=tot_obs;
  ii+1;
  if ii>tot_obs then stop;
  run;

  proc sort data=tab_boot; by nopers; run;

  data tab_boot ;
  merge tab_boot (in=ib) ent;
  by nopers;
  if ib;
  run;

  proc summary data=tab_boot nway;
  var n;
  class &dip. &gene. cspere;
  output out=titi sum=;
  run;

  proc sort data=titi ; by &dip. &gene.; run;

  proc transpose data=titi out=tata (drop = _NAME_) prefix
= cspere;
  by &dip. &gene.;
  var n;
  id cspere;
  run;

  data tata;
  T=&k.;
  set tata;
  if cspere1=. then cspere1=0;
  if cspere2=. then cspere2=0;
  if cspere3=. then cspere3=0;
  if cspere4=. then cspere4=0;
  if cspere5=. then cspere5=0;
  if cspere6=. then cspere6=0;

```

```

    if cspere7=. then cspere7=0;
    run;

    data eff;
    set eff tata;
    run;

%end;

/*transformation de la table pour avoir une matrice par bloc*/
proc sort data=eff out=eff2;
by &dip. T;
run;

data eff3;
set eff2;
%do i=0 %to &t.;
    if T=&i. then do;
        %do j=1 %to 7;
            cspere&j._&i.=cspere&j.;
        %end;
    end;
%end;
%do i=0 %to &t.;
    %do j=1 %to 7;
        if cspere&j._&i.=. then cspere&j._&i.=0 ;
    %end;
%end;
drop cspere1-cspere7;
run;

/*exporter les résultats dans un fichier .txt*/
FILENAME dat "D:\HAS4B7\kro%eval(&t.+1).txt";

DATA _null_;
SET eff3 (drop = T &dip. &gene.);
FILE dat;
PUT (_all_) (:);          * For space separated;
RUN;
%mend;

%databoot
(gene=gene3_1977,naia_inf=1948,naia_sup=1977,age=25,sexe=2,cham
p=chplc,dip=dip1c,t=100);

```

## 2) Exécuter LEM pour obtenir notre ajustement

a) Renommer le fichier KRONbtirage+1.TXT en changeant l'extension : .fre

b) Exécuter le programme LEM de la forme suivante :

```

*programme pour sortir en une seule fois les estimations de plusieurs
modèles
* cas de 101 tirages
man 3
dim 2 1010 707
lab D T    0
mod {TO, TD, fac(OD,606,T,b,910)}
dat kro2csf.fre

```

```

des
[1 2 3 4 5 6 0 7 8 9 10 11 12 0 13 14 15 16 17 18 0.....
..... 595 596 597 598 599 600 0 601 602 603 604 605 606 0
0 0 0 0 0.....0

1 2 3 4 5 6 7 8 9 10 1 11 12 13 14 15 16 17 18 19 1 ..... 875 876
877 878 879 880 881 882 883 1 884 885 886 887 888 889 890 891 892 1
893 894 895 896 897 898 899 900 901 1 902 903 904 905 906 907 908 909
910

]
nec nco nfr          Supprime des éditions dans les sorties LEM

```

### 3) Récupérer les $\beta_g^i$ , $i = 1..T + 1$ et calculer les écarts-type

a) Enregistrer la série des betas dans un fichier .txt.

b) Lancer la macro %BETABOOT qui calcule les écarts-type et mets en forme les résultats.

```

/*****
*****/
/*                                %BETABOOT                                */
/* macro permettant de mettre les 1 là ou il faut et de                    */
/* calculer les ecarts-types bootstrap*/
/* T= nombre de jeux de données (y compris celui observé)=nb            */
/* tirage+1 */
/* g= nombre de générations */
/* datain= nom du fichier txt contenant les données ie les              */
/* betas */
/*****
*****/
/*La table en sortie (CALC) contient :                                     */
/* - le numéro des générations */
/* - les betas observés */
/* - la valeur de l'Ecart type bootstrap */
/* - les betas inf et sup (=beta obs +/- 1.96*EC) */
/* - le min et le max des betas bootstrapés */
/*****
*****/

%macro betaboot(T=,g=,datain=);
/*importation du fichier txt contenant tous les betas*/
proc import out=beta DATAFILE= "D:\HAS4B7\&datain..txt"
DBMS=DLM REPLACE;
DELIMITER='09'x;
GETNAMES=NO;
DATAROW=1;
run;

/*on rajoute dans la table des betas un compteur N et on
renomme la variable en beta*/
data betab;
N=_N_;
set beta;
rename var1=beta;
run;

/*on crée le compteur qui saute de 1 tous les g-1 lignes*/
data betab2 (drop = N);

```

```

set betab;
if N<=&g. then num=N;

else do;
  select;
  %do k=2 %to &T.;
    when (N<=%eval(&g.*&k.-(&k.-1))) num=N+%eval(&k.-
1);
  %end;
end;

end;
run;

/*on crée une table b1 contenant T*g 1*/
data b1;
num=1;
do while (num<=%eval(&T.*&g.));
  beta=1;
  output;
  num=num+1;
end;
run;

/*on merge les deux tables*/
/*celle qui ne contient que les 1 à gauche, comme ça ils seront
écrasés*/
data ens;
merge b1 betab2;
by num;
run;

/*on crée une variable T contenant le n° de tirage et compt
variable allant de 1 à g puis de 1 à g...*/
data ens2;
set ens;
select;
  %do k=1 %to &T.;
    when (num<=%eval(&g.*&k.)) do;
      T=%eval(&k.-1);
      compt=num-%eval(&g.*(&k.-1));
    end;
  %end;
end;
run;

/*on transpose pour avoir les tirages en colonne*/
proc sort data=ens2 out=ens2sort; by compt;
run;

proc transpose data=ens2sort out=ens2tr (drop=_NAME_) prefix=T
;
by compt;
var beta;
id T;
run;

/*Calcul des ecarts-types et des intervalles de confiance*/
data calc (keep = compt T0 EC beta_inf beta_sup beta_min
beta_max);
set ens2tr;
Tbar=sum(of T:)/&T.;

```

```
array T T0-T%eval(&T.-1);
array E E0-E%eval(&T.-1);
do i=1 to &T.;
    E(i)=(T(i)-Tbar)**2;
end;
EC=sqrt(sum(of E:)/&T.);

beta_inf=T0-1.96*EC;
beta_sup=T0+1.96*EC;

beta_min=min(of T1-T%eval(&T.-1));
beta_max=max(of T1-T%eval(&T.-1));

run;

/*exportation des estimations dans un fichier excel*/
proc export data=calc
outfile= "filepath"
dbms=excel5 replace;
run;

%mend;

%betaboot(T=101,g=10,datain=beta101g3);
```