

# Elaboration de carrières-types représentatives des carrières complètes des salariés du secteur privé nés en 1948

*Malik KOUBI*

*INSEE, Département de l'Emploi et des Revenus d'Activité*

## Introduction

Quelle est l'ampleur de la dispersion des trajectoires salariales ? Existe-t-il des parcours salariaux plus représentés que d'autres ? La notion de carrière-type fait référence aux scénarios microéconomiques modélisant l'évolution du salaire avec l'âge, utilisés par certains organismes sociaux et d'études pour faire des prévisions et simuler des politiques économiques en matière de retraite. Les carrières-types sont donc très utiles pour l'analyse des « performances » des régimes de retraite car elles sont supposées approcher un grand nombre de carrières « réelles ». Or les travaux sur carrières-types les plus courants<sup>1</sup> prennent généralement pour référence des cas très conventionnels, comme celui d'un salarié ayant perçu sans interruption un salaire proche du SMIC, dont on ne connaît pas précisément la représentativité. Cette étude définit une méthodologie permettant d'une part d'élaborer des carrières-types sur une base statistique et d'autre part de mesurer la représentativité des carrières-types ainsi obtenues. Elle est documentée pour cela par un panel de salariés du secteur privé et semi-public nés en 1948 dont on suit les trajectoires salariales « réelles » sur plus de trente années, de 1967 à 2000. La méthodologie ainsi dégagée permet en outre de mesurer l'impact des principales caractéristiques individuelles observées sur la forme des carrières. On ne retient ici que les carrières « complètes » (définies comme celles justifiant de 20 ans de présence ou plus) et un lissage préalable est effectué, qui consiste à remplacer chaque carrière salariale individuelle par sa tendance polynomiale.

La première partie est consacrée à décrire la durée des carrières (complètes ou non) observées dans les DADS. L'élément le plus frappant est leur diversité : au même titre que les caractéristiques individuelles observables, le temps qui passe semble exacerber les différences de trajectoire entre les individus. En second lieu, on est surpris de constater que les carrières complètes ne forment qu'une minorité (certes importante) de l'ensemble des carrières observées, comme le prouvent de simples statistiques par sexe sur les durées de présence et les taux de présence par âge. La deuxième partie, en se restreignant désormais aux carrières complètes, montre des différences de carrière importantes, selon le sexe et l'âge auquel interviennent les promotions professionnelles. Ces caractéristiques ont un impact durable sur l'ensemble de la carrière. L'idée principale de la méthode employée pour élaborer des carrières types est décrite dans la troisième partie. Elle consiste à munir l'ensemble des carrières salariales d'une distance.

---

<sup>1</sup> Voir par exemple Charpin, 1999 ou le modèle de cas-types OSCARIE de la DREES.

Cette structure que l'on ajoute à l'ensemble des carrières salariales fait de celui-ci un espace géométrique, plus facile à appréhender par des méthodes statistiques. L'utilisation d'une base de polynômes orthogonaux (dits de Legendre) pour décomposer chaque carrière individuelle est l'autre ingrédient essentiel de la méthode, qui est mis en œuvre dans la quatrième partie. Ces fonctions, construites par orthonormalisation, sont des briques élémentaires qui permettent de décomposer les carrières individuelles. Chacune d'entre elles capture en effet un aspect particulier de la forme des carrières individuelles et la décomposition ainsi obtenue permet de calculer simplement la distance entre deux carrières individuelles. Cette simplification apportée à l'expression de la distance rend celle-ci implémentable par les algorithmes de classification qui aboutissent finalement à la définition des carrières-types souhaitées.

## 1. Statistiques descriptives

### 1.1. Données utilisées et champ

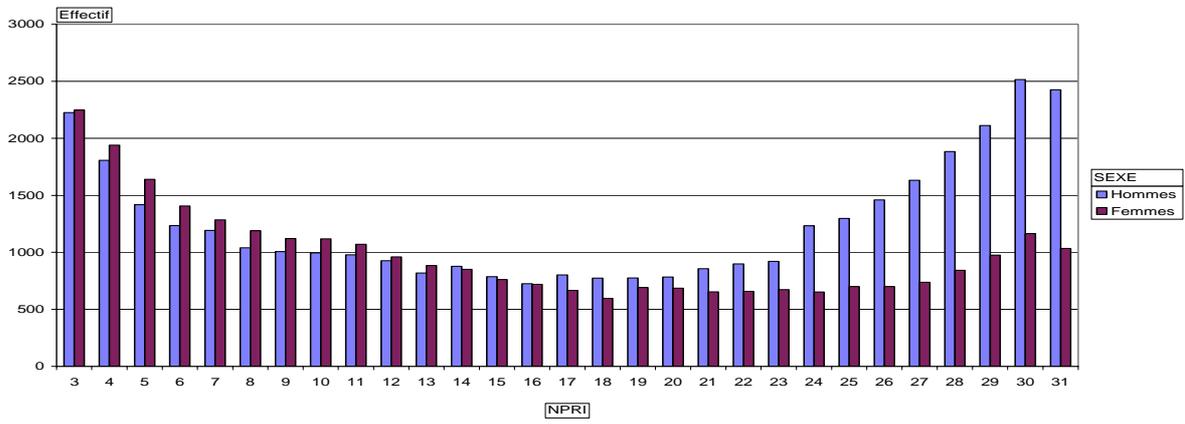
L'échantillon de salariés dont nous disposons est issu des Déclarations Annuelles de Données Sociales (DADS). Il comprend, pour chacune des années de 1967 à 2000 (hormis les années 1981, 1983 et 1990 qui ne sont pas observées), environ 4% des salariés en activité dans le secteur privé et semi-public (voir Annexe 1). Nous ne conservons que les observations relatives aux salariés nés en 1948, dont nous pouvons observer la majeure partie de la carrière. Les principales variables observées sont le salaire net, le sexe, l'âge, la durée de paie annuelle exprimée en jours, la condition d'emploi qui permet de distinguer les salariés à temps plein des formes particulières d'emploi et la catégorie socioprofessionnelle qui permet de distinguer les cadres et les non-cadres. L'échantillon total - y compris les salariés ayant eu des carrières incomplètes comporte 61706 hommes et 46738 femmes. Un grand nombre de ces salariés sont cependant observés moins de 3 ans dans le secteur privé (Tableau 1). Les autres salariés se répartissent selon le nombre d'années passées dans le secteur privé de la manière indiquée dans le graphique 2.

Les carrières très courtes (moins de 7 années de présence) et les carrières très longues (plus de 25 années) sont les plus représentées dans les DADS. Les carrières courtes correspondent vraisemblablement à des salariés n'ayant effectué que la première partie de leur carrière comme salarié du secteur privé, avant de se diriger vers la fonction publique ou de devenir indépendant. Une partie de ces salariés peut aussi s'être simplement retirée du marché de l'emploi. Si les données ne permettent pas de conclure définitivement entre ces différentes hypothèses (on ne connaît pas le motif de sortie des salariés dans les fichiers des DADS), la probabilité de présence par âge fournit cependant quelques indications utiles à ce sujet. Ainsi, l'écart de taux de participation entre les hommes et les femmes est maximal entre 25 et 32 ans (graphique 3) et le profil des taux de présence par âge est également différent entre les salariés globalement moins présents (15 ans ou moins de présence au total dans les DADS) et ceux ayant eu des carrières longues (plus de 15 ans). Le taux de présence des premiers diminue avec l'âge jusqu'à 35 ans, signe de sorties précoces du secteur privé (graphique 1). Pour les seconds, le taux de participation augmente régulièrement jusqu'à 30 ans pour les hommes et jusqu'à 40 ans pour les femmes, pour diminuer ensuite. Cela suggère que c'est en début de carrière que se décident les orientations décisives et que s'opère en particulier le choix entre une carrière dans le secteur privé après une expérience antérieure dans un autre secteur ou une carrière entièrement hors du secteur privé.

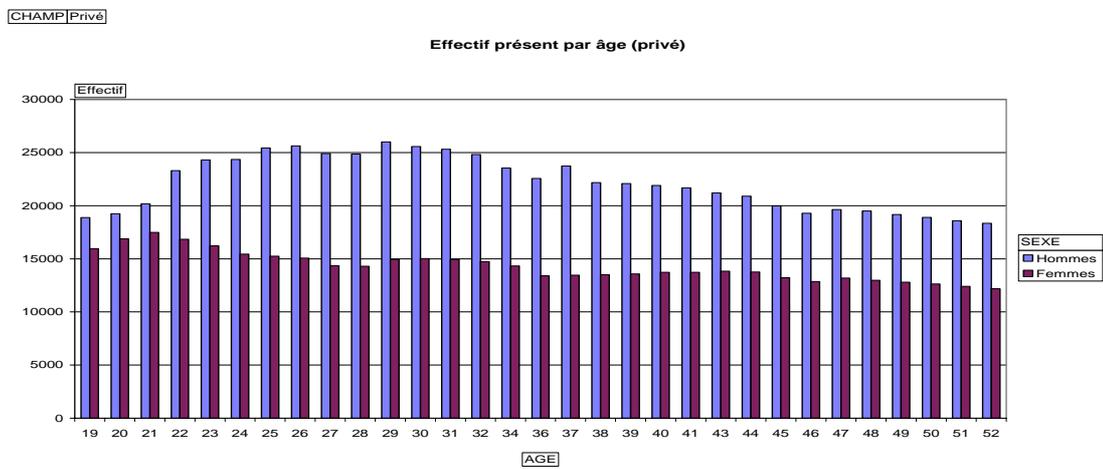
Effectif	SEXE			
	Hommes	Femmes	Total	
Nombre d'années de présence dans le secteur privé	1	19735	13120	32855
	2	4143	3287	7430
Total	23878	16407	40285	

**Tableau 1** : Nombre de salariés présents une ou deux années dans le secteur privé (la première ligne concerne les salariés du semi-public salariés présents)

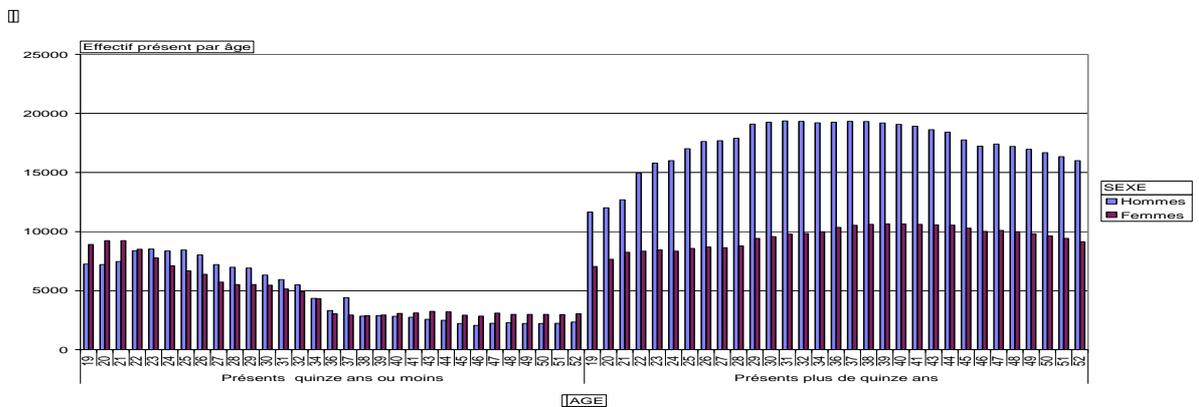
Effectif selon le nombre d'années passées dans le secteur privé



**Graphique 2 :** Répartition des salariés par nombre d'années de présence pour les salariés présents au moins trois années.



**Graphique 3 :** effectif présent dans le secteur privé, par âge.



**Graphique 4 :** Effectif présent par âge, en distinguant les salariés totalisant plus de 15 ans de présence des autres

## 1.2. Concept de salaire utilisé : annualisé, relatif

On a uniquement considéré dans cette étude des salaires nets annualisés, c'est-à-dire rapportées à la durée de paie. Le salaire annualisé est le salaire qu'aurait perçu le salarié s'il avait travaillé toute l'année au même taux de salaire journalier. Salaire perçu et salaire annualisé coïncident lorsque le salarié travaille toute l'année. C'est le cas en général sauf en début de carrière où la grande variabilité des durées travaillées annuellement confère une grande instabilité au salaire perçu. Considérer le salaire annualisé a pour effet d'atténuer le différentiel négatif très important de salaire en début de carrière.

Pour rendre comparable les salaires d'une année à l'autre, il faut d'autre part leur appliquer un déflateur annuel. Celui-ci peut être l'indice des prix à la consommation, auquel cas le salaire considéré mesure un pouvoir d'achat. On a lui préféré ici le déflateur par le salaire moyen car les carrières « relatives » (exprimé en % du salaire moyen) sont mieux expliquées que les carrières « réelles » (en francs constants) par les modèles économétriques. En effet, les salaires « relatifs » sont purgés des effets de date (chocs atteignant uniformément l'ensemble des salariés, effets conjoncturels, etc.), et les carrières « relatives » sont de ce fait plus lisses et modélisées de manière plus fiable.

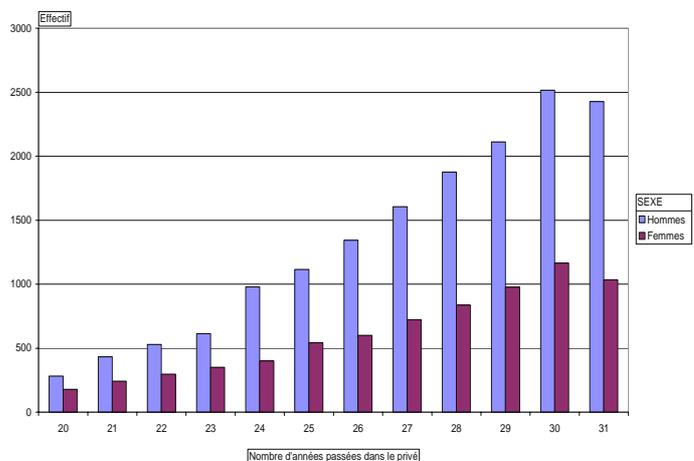
Au total, l'annualisation permet de contrôler les fluctuations de durée de paie et le déflateur par le salaire moyen celles liées aux fluctuations conjoncturelles.

## 1.3 Le champ des salariés ayant eu des carrières « complètes : une minorité des cas

On s'attache désormais au champ restreint des carrières « complètes ». Sera considérée comme complète une carrière débutant avant 30 ans, s'achevant après 45 ans, n'ayant pas d'interruption de plus de 4 années consécutives et totalisant au moins 20 années dans le secteur privé sur les 31 possibles. Un nombre de présence suffisant limite la variabilité du salaire due aux interruptions. L'échantillon ainsi restreint comporte 15831 hommes et 7345 femmes. Il est (comme on pouvait le prévoir) déséquilibré en faveur des hommes et en faveur des cadres, plus représentés parmi les salariés ayant des carrières complètes. Ces 23176 salariés ayant eu des carrières complètes totalisent 639745 années de présence, soit en moyenne 27,6 années de présence.

**Tableau 4a et Graphique 4b** : Répartition de l'échantillon des salariés ayant eu des carrières complètes, par durée de présence dans le privé.

Effectif	SEXE		
	Hommes	Femmes	Total
Nombre d'années passées dans le privé			
20	282	179	461
21	434	242	676
22	529	297	826
23	614	350	964
24	980	401	1381
25	1115	542	1657
26	1343	600	1943
27	1605	722	2327
28	1875	837	2712
29	2112	977	3089
30	2516	1164	3680
31	2426	1034	3460
<b>Total</b>	<b>15831</b>	<b>7345</b>	<b>23176</b>



## 2. Carrières « relatives » moyennes de différentes catégories de salariés

### 2.1. Le délai d'accès au statut de cadre joue influence la carrière salariale

Les carrières salariales sont influencées par le parcours professionnel. L'accès au statut de cadre marque de ce point de vue, pour les salariés qui en bénéficient, une étape importante de la carrière qui, ne se produisant pas au même âge pour tous les salariés, engendre d'importants écarts de carrière. Les salariés ont donc été classés selon le nombre d'années qu'ils ont passées sur le marché de l'emploi avant d'obtenir une promotion au statut de cadre. La typologie distingue les salariés qui ont été cadre au bout de 5 ans d'expérience ou moins, ceux qui ont été cadres après cinq années d'expérience, ceux à qui il a fallu au moins quinze ans et enfin ceux qui n'ont jamais été cadres. La répartition de l'échantillon selon ces quatre classes est donnée dans le tableau 4.

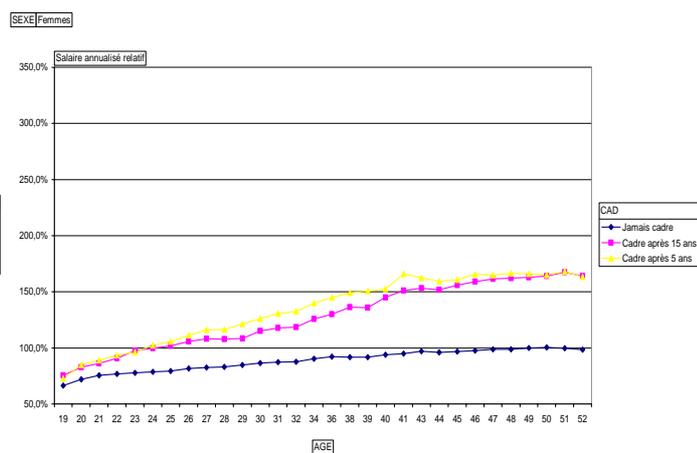
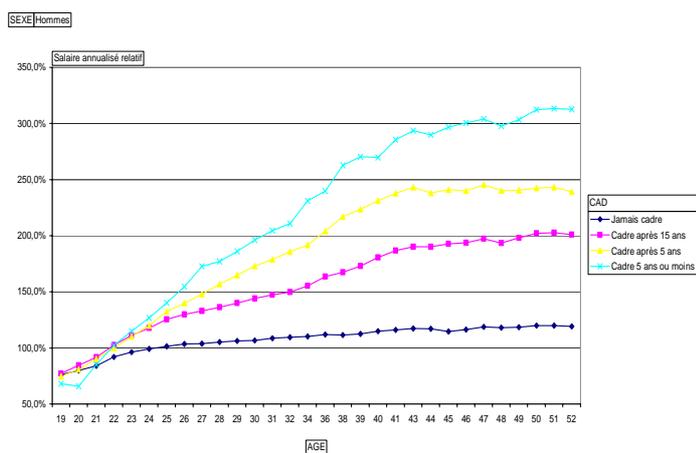
CAD	SEXE	
	Hommes	Femmes
Jamais cadre	64,2%	74,4%
Cadre après 15 ans	17,4%	14,9%
Cadre après 5 ans	13,4%	8,1%
Cadre 5 ans ou moins	5,0%	2,6%
Total	100,0%	100,0%

**Tableau 4** : Répartition de l'échantillon selon la durée d'attente avant le passage au statut de cadre

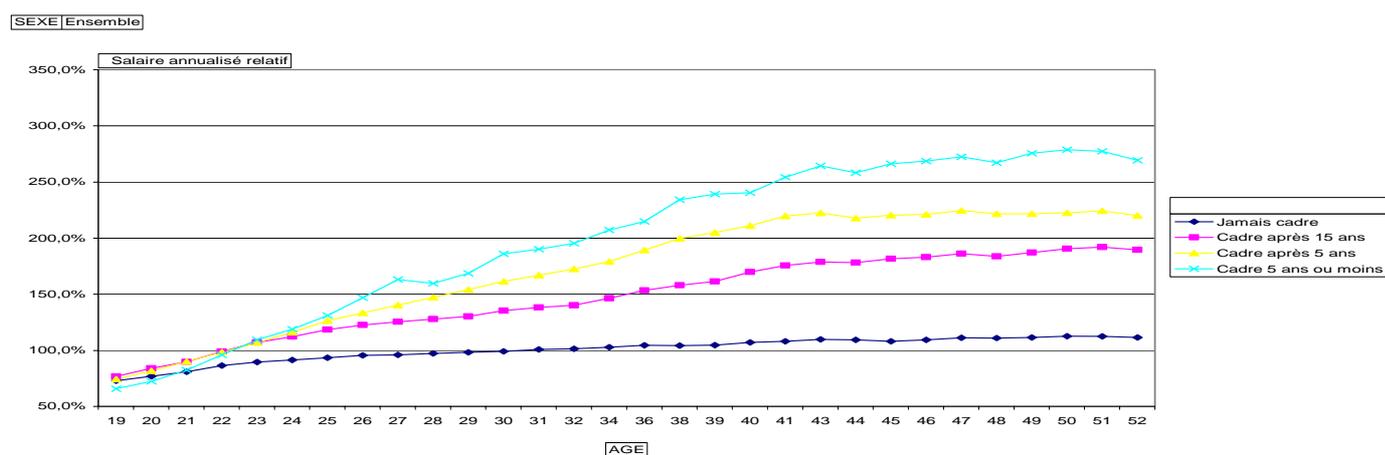
### 2.2. Carrières des cadres et des non-cadres

Les graphiques 6 et 7 présentent les carrières relatives moyennes de ces différentes catégories de salariés. En moyenne, les salariés nés en 1948 gagnaient environ 75% du salaire moyen à 20 ans (Graphique 8). Il existait alors peu d'écarts entre les futurs promus et les salariés qui resteraient non-cadres. A 50 ans, les non-cadres gagnaient 10% de plus que le salaire moyen, alors que les salariés qui avaient accédé au statut de cadre dans les cinq premières années de leur carrière gagnaient plus de deux fois et demi le salaire moyen. Les femmes nées en 1948 qui n'ont jamais été cadres gagnaient en moyenne 70% du salaire moyen à 20 ans (Graphique 7). Elles gagnaient une fois ce salaire moyen à 50 ans. Sur la même période, le salaire des hommes non-cadres est passé de 80% à 120% du salaire moyen (Graphiques 6 et 7).

L'accès au statut de cadre est plus rentable pour les hommes que pour les femmes. Ainsi, parmi les salariés qui ont accédé au statut de cadre après 15 ans de carrière, les hommes gagnent en moyenne deux fois le salaire moyen à 50 ans alors que les femmes gagnent seulement 1,6 fois ce même salaire moyen. L'âge d'accès au statut de cadre n'a par ailleurs pas seulement un effet sur le salaire permanent, mais aussi sur la progression du salaire tout au long de la vie. Le rendement de l'expérience n'évolue pas de la même manière avec l'âge d'une catégorie de salarié à l'autre. Ainsi, le salaire relatif des salariés les plus précocement promus continue de progresser après 45 ans, tandis qu'il se stabilise à partir de cet âge pour les autres catégories de salariés.



**Graphiques 6 et 7 :** Carrières relatives moyennes par sexe, pour les cadres et les non-cadres (les femmes qui ont été cadres en moins de cinq années sont trop peu nombreuses pour que leurs carrières soient représentées)



**Graphique 8 :** Carrières relatives moyennes, pour les cadres et les non-cadres

### 2.3. Quelques enseignements des résultats précédents

L'extrême diversité des carrières provient du fait que celles-ci diffèrent d'un salarié à l'autre non seulement par le niveau moyen de salaire tout au long de la carrière (le salaire permanent) mais aussi par le rendement de l'expérience et donc par la forme des carrières (forme caractérisée entre autres par le taux de croissance moyen du salaire). Une approche prenant en compte ces différentes caractéristiques de la carrière est donc nécessaire. On note aussi que les caractéristiques individuelles observables (comme le fait d'occuper un poste de cadre) sont corrélées au rendement de l'expérience et donc à la forme des carrières.

### 3. Construction de carrières-types statistiquement fondées : approche théorique

#### 3.1. Une distance pour mesurer l'hétérogénéité des carrières

La construction de carrières-types sur un critère statistique constitue l'apport méthodologique principal de cette étude. L'originalité de la méthode utilisée tient dans le fait que l'on munit l'ensemble des carrières individuelles d'une distance, quantité positive mesurant le degré de dissemblance de deux parcours salariaux individuels. La distance entre les carrières de deux salariés sera ici définie comme l'écart quadratique entre les salaires de ces deux salariés, écart évalué en moyenne tout au long de la carrière. Cette distance mesure non seulement les différences de salaire permanent entre deux carrières (le niveau moyen de celle-ci), mais aussi celles tenant à la forme globale de ces carrières (comme le taux de croissance moyen du salaire par exemple).

La structure métrique ainsi définie rend plus facile l'application des méthodes statistiques habituelles. L'ensemble des carrières doté d'une distance devient en effet un espace géométrique sur lequel peuvent s'appliquer de nombreux traitements statistiques. En particulier, chaque carrière individuelle peut dans un premier temps être décomposée à l'aide d'un ensemble de fonctions élémentaires - les polynômes orthogonaux - dont chacune capture une composante particulière de la forme de la carrière. Les carrières individuelles, une fois décomposées sur cette base de polynômes orthogonaux, peuvent être classées en utilisant les méthodes statistiques habituelles. On obtient ainsi des systèmes de carrières types de différentes tailles, statistiquement significatifs, dont la représentativité peut être mesurée a posteriori.

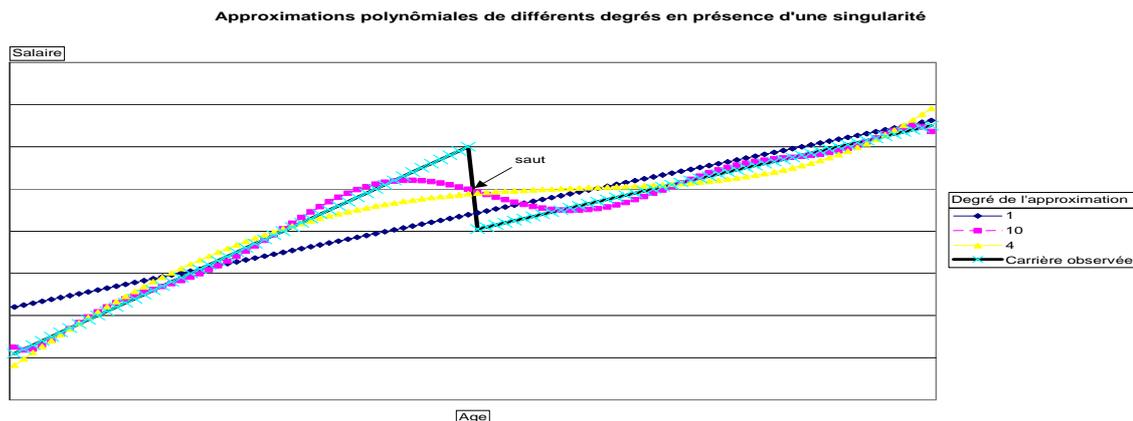
#### 3.2 Un lissage préalable des carrières individuelles

Un lissage préalable des données a été effectué, opération qui consiste à remplacer chaque carrière individuelle par sa tendance polynomiale de degré fixé. Cette opération a d'une part l'avantage de diminuer le bruit existant dans les carrières (dont on suppose qu'une partie provient d'erreurs de mesure, des années manquantes, etc.) et d'autre part de réduire le nombre de paramètres nécessaire pour représenter chaque carrière individuelle. Cette simplification a par ailleurs un prix, celui de ne pas conserver la totalité de l'information relative à la carrière que l'on cherche à approcher. Toutefois, la perte d'information observée concerne essentiellement les carrières très atypiques, qui s'éloignent de manière sensible de leur tendance. Elle peut de plus être réduite en considérant des approximations polynomiales de degré suffisamment élevé. Le graphique 9 simule en guise d'illustration le comportement des approximations de différents degrés au voisinage d'une discontinuité (un « saut ») de la carrière salariale. Notons aussi que même si l'approximation polynomiale peut s'éloigner de la valeur réelle du salaire à tel ou tel âge, elle n'en conserve pas moins les principales caractéristiques de la carrière, comme la moyenne des salaires sur toute la carrière (le salaire permanent)<sup>2</sup>.

In fine, le choix du degré de l'approximation résulte de manière classique d'un compromis entre la précision désirée dans l'approximation des carrières et la parcimonie des paramètres descriptifs pour les représenter. Ce choix s'est porté sur des polynômes de degré 4, qui ont le mérite de bien approcher la plupart des carrières sans perte trop importante d'information, comme l'atteste le calcul des coefficients de détermination ( $R^2$ ) des régressions effectuées.

---

<sup>2</sup> Cette propriété peut par exemple être déduite du fait que le polynôme constant appartenant au sous-espace des approximations polynomiales que l'on considère, il est orthogonal à la différence entre la carrière  $f$  et son approximation  $\hat{f}$ , de sorte que finalement  $E_e(f) = E_e(\hat{f})$  : la carrière observée et son approximation ont le même salaire permanent (le même niveau moyen).



**Graphique 9 :** Approximations polynômiales de différents degrés. La courbe ayant un saut représente la carrière observée d'un salarié. Ce saut marque un changement important dans la carrière, à la suite duquel le niveau de salaire et la pente de la carrière changent. Les courbes lisses représentent quant à elles des approximations polynômiales de différents degrés.

### 3.3 Comment « métriser » la diversité des carrières individuelles

Même après cette première simplification, la diversité des carrières individuelles reste très importante. En effet, les disparités de salaire que l'on observe une année donnée sont amplifiées par les différences de parcours entre salariés, différences qui s'accroissent au cours du temps. Les salariés voient ainsi leurs trajectoires diverger au gré des décisions individuelles ou sous l'effet d'événements divers qui jalonnent leurs parcours et marquent leur devenir professionnel.

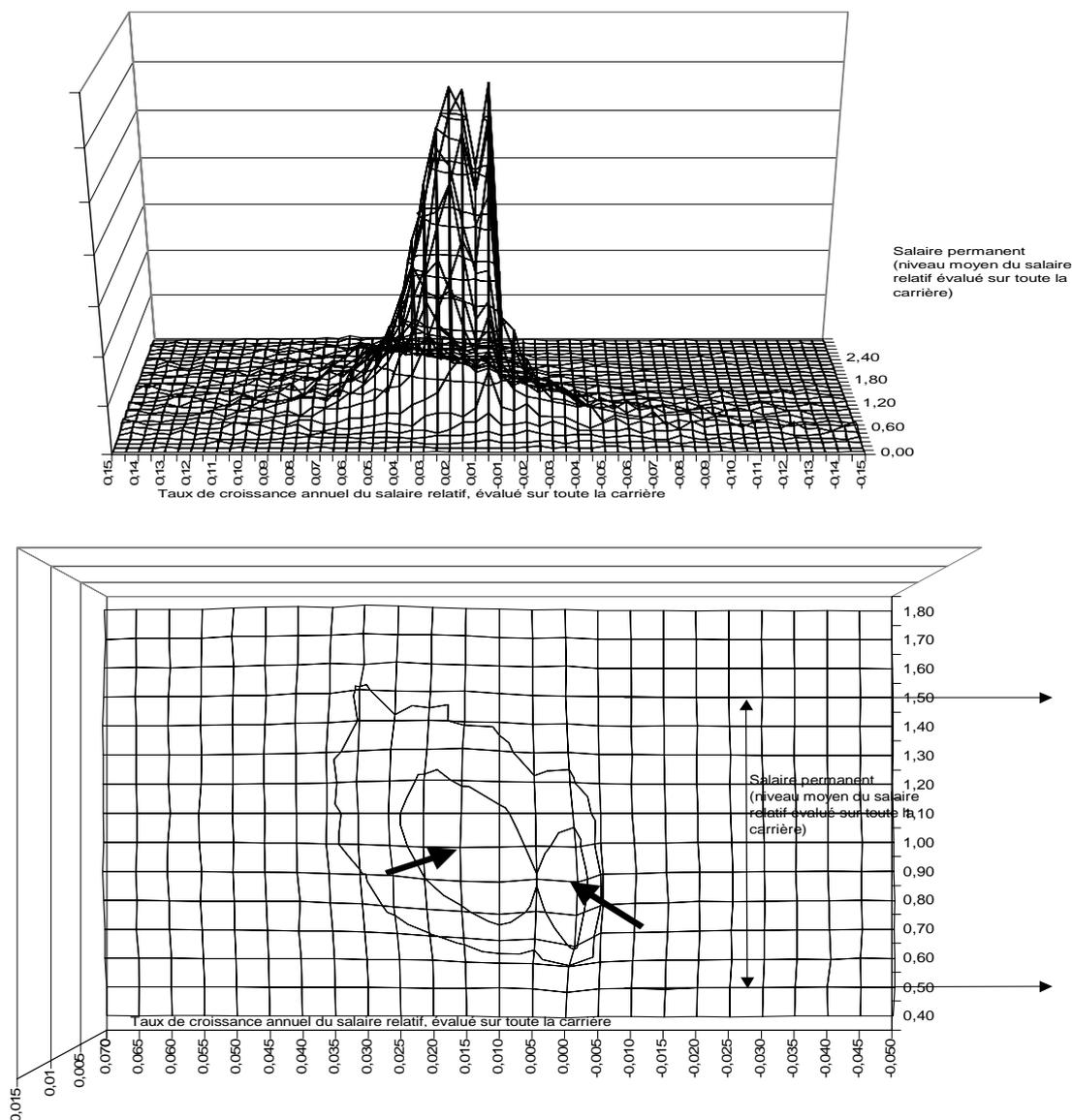
Le graphique 10 présente une vue partielle de la distribution des carrières individuelles, dans un plan dont le premier axe est constitué par le salaire permanent (salaire évalué en moyenne sur toute la carrière) et le second axe le taux de croissance annuel du salaire, évalué lui aussi en moyenne sur la carrière. Salaire permanent et taux de croissance du salaire sont positivement corrélés entre tous les salariés et la très grande majorité des carrières possède un salaire permanent compris entre 0,5 et 1,5 fois le salaire moyen et un taux de croissance compris entre -0,5% et 3,5% par an. Cette règle générale n'exclut toutefois pas des exceptions notables, car la densité des carrières reste différente de zéro pour un salaire trois fois supérieur au salaire moyen et un taux de croissance du salaire en deçà de 5% et au-delà de 10%.

Les carrières salariales forment un ensemble dont la complexité est telle qu'elle rend impossible toute approche directe. Construire une typologie des carrières salariales passe donc certainement par des méthodes statistiques adaptées mais suppose aussi d'avoir préalablement défini une structure supplémentaire qui permette d'appliquer ces méthodes. Cette structure prend ici la forme d'une distance, indicateur qui mesure le degré de dissemblance (ou de similitude) entre deux carrières salariales réelles. Elle est définie comme l'écart quadratique entre les salaires correspondant aux deux profils salariaux, écart évalué en moyenne sur toute la carrière. Cette distance prend en compte non seulement les différences de salaire permanent, mais aussi les différences de taux de croissance moyen etc.... Ainsi, la distance entre les carrières de deux salariés n'est nulle que s'ils ont perçu les mêmes salaires aux mêmes dates. Elle n'est en particulier pas nulle s'ils ont perçu le même salaire en moyenne sur toute la carrière, mais que leur carrière diffère seulement par la répartition de leur revenu tout au long de la carrière. Quelques exemples de cette distance sont donnés dans le tableau 17 et le graphique 18 (page 15).

$$distance(f, g)^2 = E_e \left( (f - g)^2 \right) = \frac{1}{T} \sum_1^T (f_t - g_t)^2$$

La distance ainsi définie possède en outre une autre propriété importante qui la rend utilisable - moyennant quelques manipulations techniques - par les algorithmes de classification habituels.

En effet, cette distance définit sur l'ensemble des carrières salariales une structure d'espace vectoriel « euclidien ». Dans un tel espace, on peut construire des fonctions particulières (les bases de polynômes orthogonaux) qui permettent de décomposer de manière simple et efficace chaque carrière réelle. Lorsque les carrières individuelles sont décomposées dans cette base, l'expression de la distance prend une forme particulièrement simple, forme directement utilisable par les algorithmes de classification habituels. La construction de cette base de polynômes orthogonaux et les simplifications que ceux-ci amènent dans l'expression de la distance sont expliquées dans l'annexe 2.



**Graphique 10 :** Distribution des carrières complètes dans le repère (revenu permanent, taux de croissance du salaire) : ces deux graphiques sont deux vues de la même surface, qui donne la répartition des salariés selon leur salaire permanent (le niveau moyen de leur salaire au cours de leur carrière) et le taux de croissance annuel moyen de leur salaire. L'axe des abscisses indique la pente de la carrière (en taux de croissance annuel du salaire) et l'ordonnée le niveau moyen du salaire au cours de la carrière (en % du salaire moyen de l'ensemble de l'économie). La distribution est bi-modale : la carrière la plus représentée est en moyenne au niveau du salaire moyen et croît de 1,5% par an. Le second mode se situe autour de 0,9 fois le salaire moyen et correspond à une croissance quasiment nulle du salaire (en salaire nominal, cela signifie une croissance au rythme du salaire moyen). Les courbes délimitent des régions de densités différentes : les régions au plus centre sont celles où le nombre de salariés est le plus important, et la densité décroît quand on va vers les régions extérieures.

### 3.4 Une définition géométrique des carrières-types

Un Système de carrières-types peut être défini de manière intuitive comme un ensemble de carrières salariales, réelles ou fictives, vérifiant le double critère de représentativité et de parcimonie suivant : d'une part toute carrière réelle doit être assez proche d'au moins une des carrières types du Système et d'autre part le nombre des carrières-types du Système doit être le plus petit possible. Dans l'espace des carrières désormais muni d'une distance, cette définition peut être reformulée dans le langage de la géométrie, en faisant appel à la notion de « réseau », qui formalise la propriété de recouvrement d'un espace métrique à l'aide d'un petit nombre de points (voir encadré 1 et figure 11).

#### Encadré 1 : La notion de carrière type traduite dans le langage de la géométrie

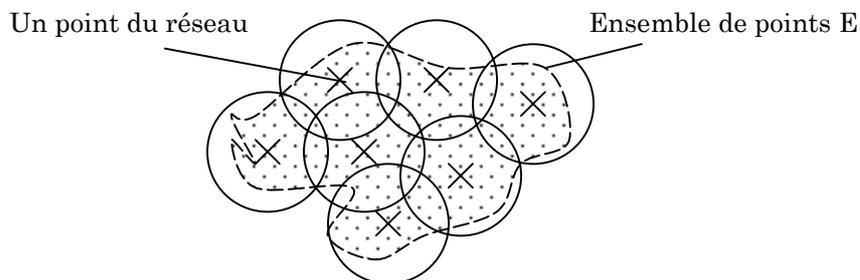
La propriété de recouvrement d'un espace métrique à l'aide d'un petit nombre de points, qui s'applique aux carrières types, est connue en géométrie sous la définition de « réseau ». Plus précisément, pour une valeur positive du paramètre  $\varepsilon$ , un  $\varepsilon$ -réseau est un ensemble de points, le plus petit possible, qui « recouvre » avec une erreur au plus égale à  $\varepsilon$  l'ensemble des points de E. Autrement dit, tout point de l'ensemble est situé à une distance inférieure à  $\varepsilon$  d'un des points du réseau. Un système de carrières types peut être défini comme un réseau de l'ensemble des carrières. Dans cette définition, la valeur  $\square$  est un degré de précision (mesuré en % du salaire), posé a priori, avec lequel on souhaite que le Système de carrières types approche l'ensemble des carrières réelles. Ou encore, c'est encore l'erreur maximale que l'on commet lorsqu'on remplace une carrière réelle donnée par la carrière type la plus proche. Cette propriété de recouvrement par des boules de rayon centrées autour des carrières-types est illustrée par la figure 11.

De manière formelle, les  $\varepsilon$ -reseaux dans l'ensemble E sont en toute rigueur caractérisés par une condition d'optimalité non linéaire et difficile à implémenter numériquement. On définit d'abord, pour un ensemble de points  $R = (r_1, \dots, r_n)$ , la fonction

$$\varphi(R) = \text{Max}_{x \in E} \text{Min}_{r \in R} (\text{dist}(x, r)) = \text{Max}_{x \in E} (\text{dist}(x, R))$$

Cette quantité représente la précision avec laquelle le n-uplet  $R = (r_1, \dots, r_n)$  de carrières approche l'ensemble des carrières. Les  $\varepsilon$ -reseaux peuvent alors être définis comme les ensembles de points  $R = (r_1, \dots, r_n)$  pour lesquels la fonction d'approximation est inférieure à  $\varepsilon$  et qui sont de plus de cardinalité minimale. Ils réalisent ainsi le minimum sous contraintes du programme suivant :

$$\text{Min}_{\varphi(R) \leq \varepsilon} (n) = \text{Min}_{\varphi(R) \leq \varepsilon} \text{Cardinal}(R)$$



**Figure 11a :** Dans l'ensemble E, assimilable à l'ensemble des carrières salariales, a été représenté un ensemble de 7 points formant un  $\varepsilon$ -réseau, repérables par une croix. Cet ensemble de 7 points est assimilable à un Système de carrières-types. Les boules qu'on a représentées autour de ces points et dont ils sont le centre ont chacune pour rayon  $\varepsilon$  : elles représentent l'ensemble des carrières qui peuvent être approchées avec une erreur au plus égale à  $\varepsilon$  par la carrière-type située en son centre. Dans le cas représenté, ces boules recouvrent tout l'ensemble E, ce qui signifie que tout point de l'ensemble E est à une distance inférieure à  $\varepsilon$  d'un des points du réseau.

## 4. Construction de carrières-types statistiquement fondées : mise en pratique

### 4.1. Une construction des carrières-types par classification

Si la notion géométrique de réseau caractérise en toute rigueur l'objet recherché (les carrières types), cette définition reste toutefois difficile à implémenter numériquement, le problème à résoudre étant fortement non linéaire. En revanche, cette définition suggère d'approcher l'objectif fixé, qui est de construire des carrières représentatives de la manière la plus économique, à l'aide des méthodes -plus classiques - de classification. Ces algorithmes permettent en effet d'obtenir un ensemble de carrières répondant le mieux à la définition d'un Système de carrières types. Les algorithmes de classification aboutissent en effet à la définition de classes d'objets en regroupant de proche en proche les objets les moins distants. En conséquence, les classes obtenues ont la propriété que deux objets appartenant à une même classe sont proches et deux objets appartenant à des classes différentes sont éloignés. Les centres des classes obtenues constituent ainsi un ensemble de points aux propriétés proches de celle d'un réseau.

En pratique les Systèmes de carrières types seront obtenus par classification des carrières sur le critère de la distance que l'on a définie. Ce type de méthodes statistiques consiste à regrouper entre elles les carrières proches de façon à obtenir des classes homogènes. La carrière moyenne de chaque classe est ainsi représentative de l'ensemble des carrières de la classe.

A ce stade, il subsiste toutefois un obstacle, qui réside dans la forme de la distance utilisée. Celle-ci en effet, égale à l'écart quadratique moyen des carrières, n'est en effet pas utilisable en l'état par les logiciels de statistique et doit donc être préalablement simplifiée.

### 4.2. Des polynômes orthogonaux pour décomposer les carrières individuelles

La simplification de la distance passe par la décomposition de chaque carrière individuelle dans une base de fonctions élémentaires : les polynômes orthogonaux. Ces fonctions, construites par orthonormalisation, sont des briques élémentaires qui permettent de décomposer les carrières individuelles. Chacune de ces fonctions capture en effet une composante particulière de la forme des carrières individuelles, composante qui est de plus orthogonale à celles capturées par les autres polynômes orthogonaux. La famille de polynômes orthogonaux utilisée ici est décrite dans l'annexe 3.

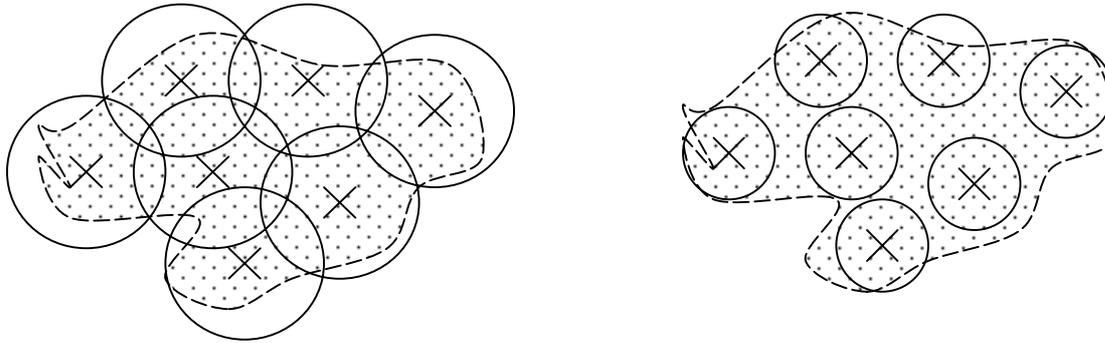
La décomposition ainsi obtenue permet ainsi ex-post de calculer simplement la distance entre deux carrières individuelles. Cette simplification apportée à l'expression de la distance rend celle-ci implémentable par les algorithmes de classification qui aboutissent à la définition de classes homogènes de carrières et ainsi des carrières-types souhaitées (voir annexe 2).

### 4.3. Mesure de la représentativité des carrières types

La représentativité d'un Système de carrières types peut par ailleurs être évaluée en calculant, pour tout nombre positif  $\varepsilon$ , la proportion de carrières situées à une distance inférieure à  $\varepsilon$  d'une des carrières-types du Système (c'est la mesure du  $\varepsilon$ -voisinage du Système). L'écart quadratique étant mesuré en logarithme du salaire, la distance  $\varepsilon$  s'exprime en %. On obtient ainsi une courbe de précision pour chaque Système de carrières types. En faisant varier la valeur de  $n$  et de  $\varepsilon$ , on obtient un ensemble de courbes donnant, pour chacune des valeurs du couple  $(n, \varepsilon)$ , la proportion de carrières  $p(n, \varepsilon)$  représentable avec une précision (ou une erreur maximale) égale à  $\varepsilon$  par un Système de  $n$  carrières-types. (graphiques 12 et 13).

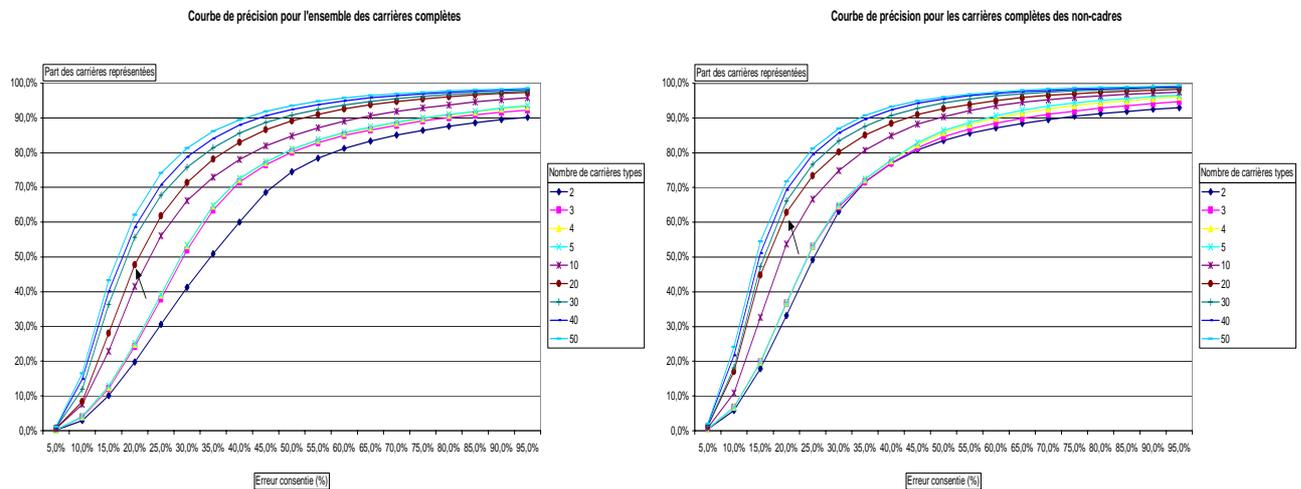
$$P(n = 7, \varepsilon = 50\%) = 1$$

$$P(n = 7, \varepsilon = 25\%) = 0,4$$



**Figure 11b** : La taille des cercles indique l'erreur maximale commise lorsqu'on approche l'ensemble des carrières réelles par le Système de carrières types. Avec une erreur maximale de 50%, le Système de 7 carrières types parvient à représenter 100% des carrières réelles. Si on est plus exigeant (erreur maximale de 25%), la proportion de carrières représentées est de 40%.

La proportion de carrières représentées lorsqu'on s'autorise une erreur très grande ( $\varepsilon \rightarrow \infty$ ) ou un nombre de carrières types très important, tend vers 1. Il n'est pas rentable d'aller au-delà d'un jeu de 20 carrières types car au-delà de ce nombre, ajouter une carrière type supplémentaire dans le Système ne permet pas d'accroître suffisamment la proportion de carrières réelles représentées. Il subsiste alors une proportion située entre 20% et 30% de carrières trop atypiques pour être représentées avec parcimonie.



**Graphiques 12 et 13.** Lecture (flèche) : Avec un jeu de 20 carrières-types on peut représenter près de 50% de l'ensemble des carrières des salariés nés en 1948 ayant eu des carrières complètes, et ce avec une précision de 20%. Pour les mêmes paramètres on parvient à représenter une proportion de plus de 60% des carrières des non-cadres,

#### 4.4. Description des carrières types obtenues

Intéressons-nous désormais à la description des carrières types obtenues, en se restreignant au jeu présenté dans la figure 14 : il s'agit des trois carrières types obtenues par classification des carrières des salariés qui n'atteignent pas le statut de cadre. La première carrière est haute et croissante : elle débute à 0,8 salaire moyen à l'âge de 19 ans pour atteindre 1,6 fois ce salaire moyen à 52 ans. La deuxième carrière est moyenne est légèrement croissante (de 0,75 salaire moyen à 1 salaire moyen). La troisième est basse et plate sauf en début et en fin de carrière où le niveau du salaire est plus bas. Son niveau ne s'élève jamais au-dessus de 60% du salaire moyen.

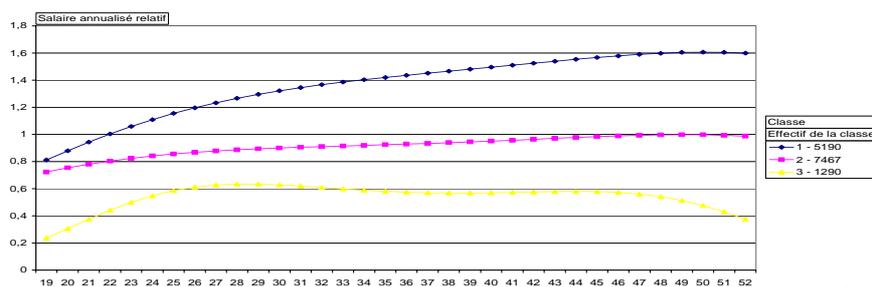
En dehors de la suite des salaires correspondant à ces carrières, il est utile de décrire au moyen de leurs caractéristiques individuelles observables les salariés dont la carrière se ramène à chacune de ces carrières. Dans ce but on a calculé, dans chacune des classes obtenues, la part représentée par plusieurs catégories de salariés : part des femmes, des salariés non à temps plein, des professions intermédiaires, des ouvriers et des salariés travaillant en région parisienne. Comme par ailleurs certaines caractéristiques individuelles changent au cours du temps, en raison de la mobilité des salariés, la structure des différentes classes de salariés a été calculée à

chaque âge, ce qui permet d'inclure des aspects dynamiques dans la description des différentes classes de salariés. Ces statistiques font l'objet des graphiques 16 à 20. Notons que même pour les caractéristiques fixes comme le sexe, il peut y avoir des variations de la structure des classes au cours du temps en raison des entrées et des sorties des salariés. En effet tous les salariés considérés ne sont pas présents à tous les âges et les statistiques concernant un âge donné sont faites en prenant pour base les salariés présents à cet âge.

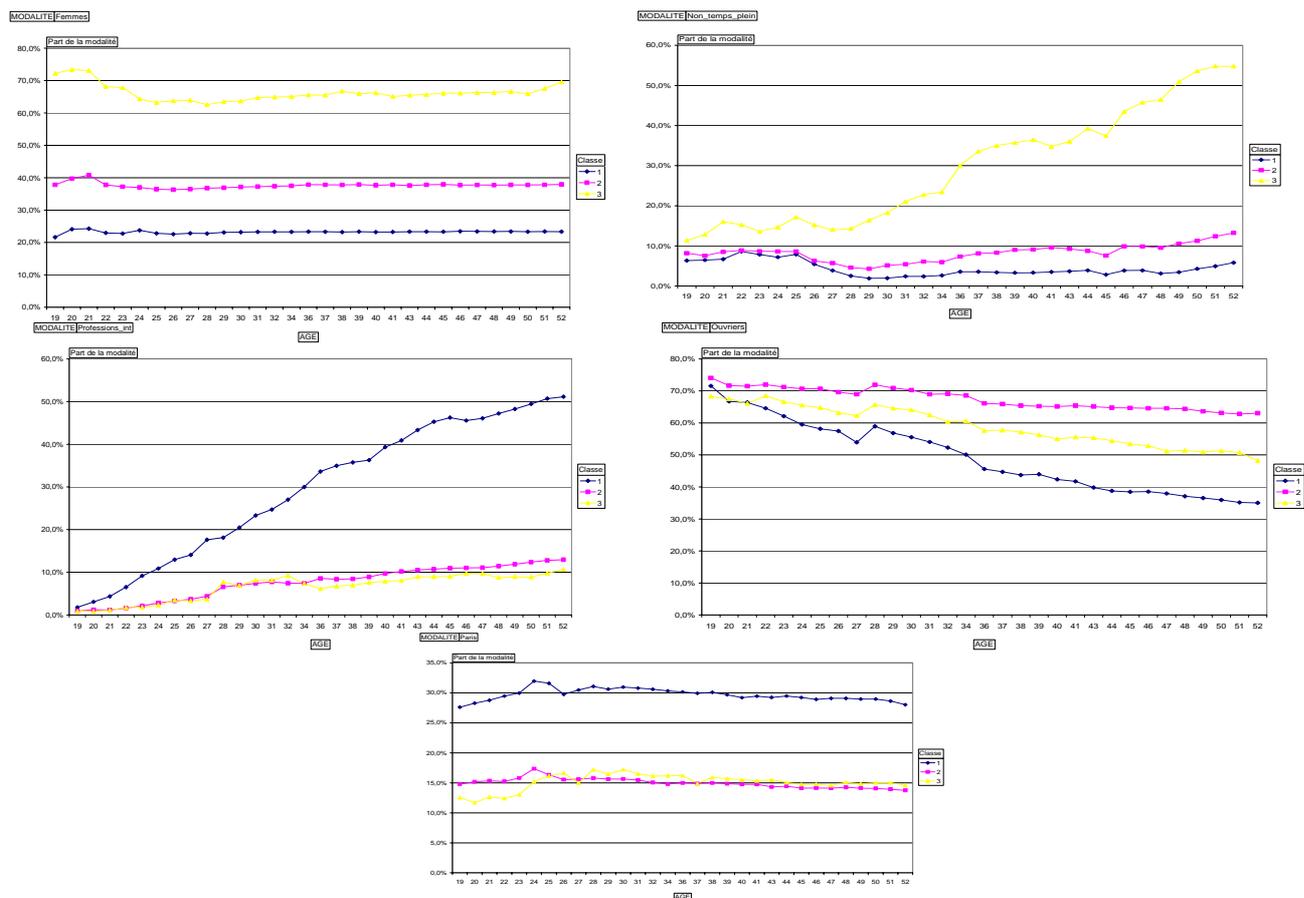
C'est parmi les salariés de la première classe de carrières (la plus haute et la plus ascendante) que les promotions vers les « Professions intermédiaires » sont les plus importantes. A l'inverse, dans la troisième classe se multiplient les passages vers les postes non à temps plein. Les trois classes de carrières sont par ailleurs ordonnées de la plus haute à la plus basse par la proportion de femmes et la première classe est caractérisée par une proportion plus importante de salariés exerçant leur activité dans la région parisienne.

[NCL3.00]

Un jeu de 3 carrières-types pour les non-cadres ayant eu des carrières complètes



**Graphiques 14:** Systèmes de trois carrières-types obtenus par classification des carrières des non-cadres.



**Graphiques 16a à 16e:** Structure âge par âge des classes de salariés formant le Système de carrières types à trois classes pour les non-cadres.

Lecture (graphique en haut à droite) : un peu moins de 20% des salariés de la classe 3 n'exercent pas leur activité à temps plein à 30 ans. A 50 ans, cette proportion dépasse les 50%.

#### 4.5. Prendre en compte les années d'activité avec des pondérations différentes

Cette étude, dont le but est essentiellement méthodologique, a pris comme illustration la description des carrières salariales. La méthode que l'on présente est toutefois assez souple pour s'appliquer à d'autres problématiques plus spécifiques. Un débouché naturel consisterait par exemple à appliquer la méthode de façon à « coller » le mieux possible au calcul des taux de remplacement associés aux régimes de retraite.

Plus précisément, il s'agit de modifier la définition de la distance de façon à ce que celle-ci prenne plus spécifiquement en compte la partie de carrière qui sert au calcul des pensions. Pour la nouvelle distance, les carrières de deux salariés doivent être proches lorsque les pensions qui leur seront versées seront proches, ce qui suppose de pondérer différemment les différentes années d'activité, voire de retenir seulement une partie d'entre elles. La prise en compte différenciée des différentes années est formalisée dans la définition de la distance par l'affectation d'un poids  $p(t)$  à chaque année  $t$ , ce qui conduit à la définition suivante de la nouvelle distance :

$$distance_p(f, g)^2 = E_e^p((f - g)^2) = \frac{1}{T} \sum_1^T p(t) \cdot (f_t - g_t)^2$$

où  $p(t)$  sont des poids vérifiant :

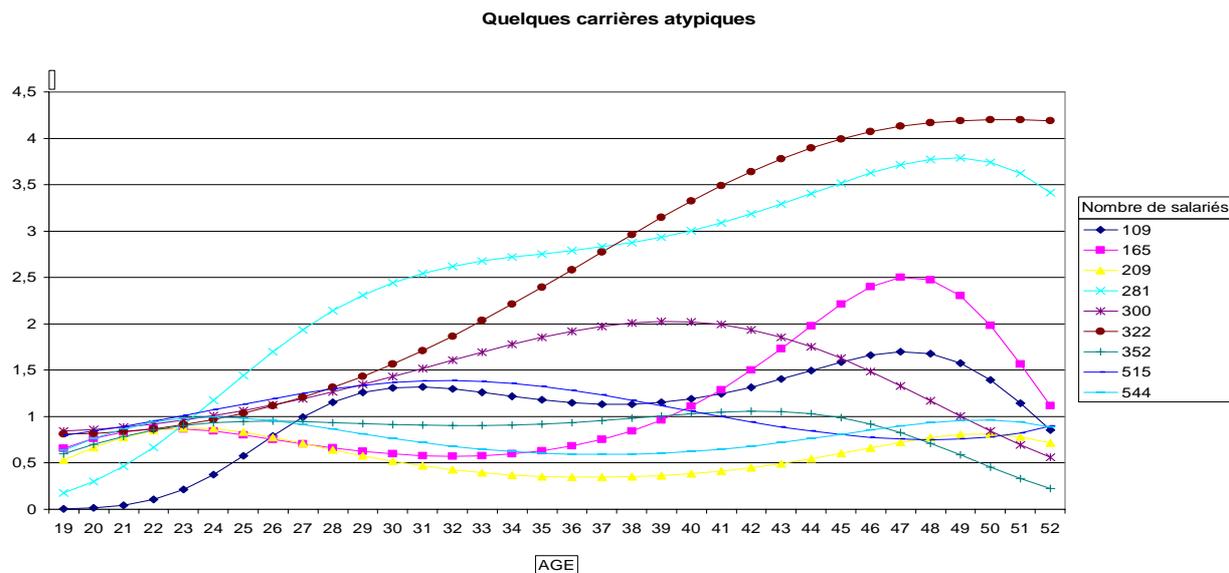
$$\sum_1^T p(t) = 1$$

Par exemple, pour tenir compte de la réforme Balladur de 1993 (où seules les 25 meilleures années sont prises en compte), on ordonnerait la séquence des salaires perçus par chaque salarié dans un ordre croissant (et non plus dans un ordre chronologique) et on affecterait un poids  $p(t) = 1/25$  pour les 25 dernières années et  $p(t) = 0$  pour les autres années.

En pratique, le procédé de décomposition des carrières s'effectue quant à lui mutatis mutandis en remplaçant chaque carrière individuelle  $f_t$  par la carrière  $f'_t = f_t \sqrt{p(t)}$ . Cela modifie en particulier la base de polynômes orthogonaux que l'on utilise mais ne change pas la méthode utilisée (voir annexe 2).

#### 4.6 Quelques carrières atypiques

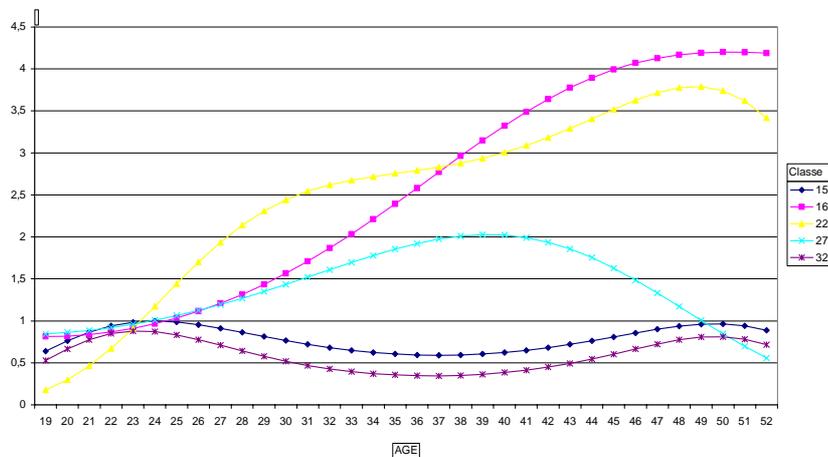
Pour illustrer le fait qu'il existe dans le secteur privé des salariés ayant suivi des carrières peu régulières, on a sélectionné, parmi le jeu de carrières du Système de 50 carrières types, quelques carrières faiblement représentées (Graphique 17). On a indiqué l'effectif des classes correspondant à chacune des carrières.



**Graphique 17 : Quelques carrières réelles mais atypiques**

Le calcul de la distance entre certaines de ces carrières illustre par ailleurs l'importante diversité des carrières. Le tableau 17 ordonne les couples de carrières par distance croissante. Ainsi, les deux plus proches carrières sont les carrières numérotées 15 et 32, puis vient le couple (16,22) etc....

Classe 1	Classe 2	Distance
15	32	0,42
16	22	0,50
15	27	1,07
16	27	1,13
22	27	1,24
27	32	1,89
15	16	2,36
15	22	2,54
16	32	3,63
22	32	3,87



**Tableau 17 et Graphique 18 : Distances entre quelques carrières atypiques**

## Bibliographie

- [1] Bayet A. (1996), « Carrières continues, carrières incomplètes et salaires », *Economie et statistique*, n° 299, pp. 21-36.
- [2] Becker G.S. (1975), *Human capital: a Theoretical and Empirical Analysis*, 2ème édition, Columbia University Press, New York.
- [3] Charpin J.M. (1999), « L'avenir de nos retraites », rapport au Premier ministre, la Documentation française, Paris 1999.
- [4] Conseil d'Orientation des Retraites (2001), « Renouveler le contrat entre les générations », premier rapport, 2001.
- [5] Guillotin Y.(1982), « Les carrières salariales en France de 1967 à 1982 », *Economie et statistique*,
- [6] Guillotin Y., Sevestre P. (1994), « Estimations de fonctions de gains sur données de panel endogénéité du capital humain et effets de la sélection », *Economie et Prévision*, n° 116, pp. 119-135.
- [7] Kapteyn A., Wansbeek T., (1989) “Estimation of the error components model with incomplete Panels”. *Journal of Econometrics* 41, 341-361.
- [8] Koubi M.(2002), « Eléments de Caractérisation des Carrières salariales des générations nées entre 1908 et 1980, Document de travail INSEE F0205, Septembre 2002.
- [9] Koubi M. Roux S.(2004), « Refonte du Panel DADS : principes et premières estimations d'emploi et de salaire », Document de travail INSEE, à paraître.
- [10] Le Minez S., Roux S.(2001), « Les différences de carrière salariale à partir du premier emploi », *Economie et Statistique* n°351, Août 2001.
- [11] Lhéritier J.L. (1992), « Les déterminants du salaire », *Economie et statistique* n°257, Septembre 1992, pp. 9-22.
- [12] Lollivier S., Payen J.F. (1990), « L'hétérogénéité des carrières individuelles mesurée sur données de panel », *Economie et Prévision*, n° 92-93, pp. 87-96..
- [13] Nijman T., Verbeek M. (1992), « Incomplete Panels and Selection Bias », *The Econometrics of Panel Data*, L. Matyas and P. Sevestre eds., Kluwer.

## Annexes

### Annexe 1 : Champ et forme des données

#### La source DADS et les filtrages

La Déclaration Annuelle des Données Sociales est une formalité déclarative que doit accomplir toute entreprise employant des salariés, en application de l'article R243-14 du code de la Sécurité Sociale (Décret du 24 mars 1972) et des articles 87.240 et 241 de la loi 51-711 du 7 juin 1951 du code Général des Impôts. Dans ce document commun aux administrations sociales et fiscales, les employeurs, y compris les entreprises nationales, les administrations publiques<sup>1</sup> et les collectivités locales, sont tenus de communiquer annuellement, pour chaque établissement, aux organismes de Sécurité Sociale d'une part, à l'administration fiscale d'autre part, la masse des traitements qu'ils ont versés (y compris primes d'ancienneté, heures supplémentaires, majorations de travail de nuit, avantages en nature, etc.), les effectifs employés et une liste nominative de leurs salariés indiquant pour chacun le montant perçu des rémunérations salariales. C'est grâce à la Déclaration Annuelle de Données Sociales qu'une telle transmission d'informations s'opère. Ce sont chaque année quelques 30 millions de déclarations « salariés » correspondant à 1,8 millions d'établissements « employeurs » qui sont ainsi enregistrées.

En 2000, le champ des DADS recouvre l'essentiel des secteurs privé et semi-public. Il représente plus de 75% des emplois salariés. Seuls les agents des organismes de l'Etat (titulaires ou non), les salariés de l'agriculture et de la sylviculture et les salariés du secteur des services domestiques ne sont pas couverts. Enfin, les collectivités territoriales ne sont couvertes entièrement par les DADS que depuis 1987. L'échantillon au 1/25 ème des DADS dont on dispose couvre les années 1967 à 2000. On n'a en outre retenu chaque année que les salariés âgés de 15 à 65 ans ayant travaillé plus de 30 jours dans l'année.

On a éliminé systématiquement certains secteurs d'activité : secteur des activités extraterritoriales, de la pêche (seul secteur agricole pouvant être présent dans les DADS).

On n'a retenu que les salariés ayant effectué l'essentiel de leur carrière dans le secteur privé. On a cependant retenu pour ceux-ci les années effectuées dans le secteur semi-public. Le champ du secteur privé est reconstitué en tenant compte au mieux des frontières. En particulier ont été considérés comme relevant du « semi-public » les postes :

- des collectivités territoriales (titulaires ou non) ;
- des grandes entreprises nationales, sauf ceux de la Poste et de France Télécom ;
- des entreprises ayant fonction d'exploitant public, des établissements hospitaliers, des établissements publics locaux sociaux et médico-sociaux, des établissements publics nationaux à compétence territoriale limitée et de ceux à caractère administratif.

Ce champ a été homogénéisé dans le temps en considérant qu'un établissement restait dans le même secteur institutionnel sur l'ensemble de la période. Chaque établissement s'est ainsi vu attribuer son secteur d'appartenance à la date la plus récente, ce qui permet d'éviter les flux d'emploi trop importants consécutifs au changement de statut de certaines grandes entreprises nationales.

## Annexe 2 : Construction d'une distance euclidienne à l'aide de polynômes orthogonaux

Dans cette étude, l'espace des carrières individuelles a été projeté sur l'espace plus petit des tendances polynomiales de degré 4, en remplaçant chaque carrière individuelle par sa tendance temporelle polynomiale de degré 4. Cet espace est naturellement muni d'une structure d'espace vectoriel de dimension cinq. Il peut être muni d'une structure supplémentaire, celle d'un espace euclidien. L'intérêt de cette structure est que la distance entre deux carrières salariales se calcule de manière très classique et qu'elle est utilisable par les algorithmes statistiques habituels. On peut alors obtenir les carrières-types par une classification hiérarchique ascendante qui utilise cette distance. Cette annexe présente la méthode, basée sur l'utilisation de polynômes orthogonaux, qui permet de construire cette distance.

### Une distance discriminante mesurant l'écart moyen de salaire entre deux carrières

Sur, l'espace des tendances polynomiales, on définit la distance entre deux tendances comme l'écart quadratique entre ces deux tendances, évalué en moyenne sur toute la carrière.

$$distance(f, g)^2 = E_e((f - g)^2) = \frac{1}{T} \sum_1^T (f_t - g_t)^2$$

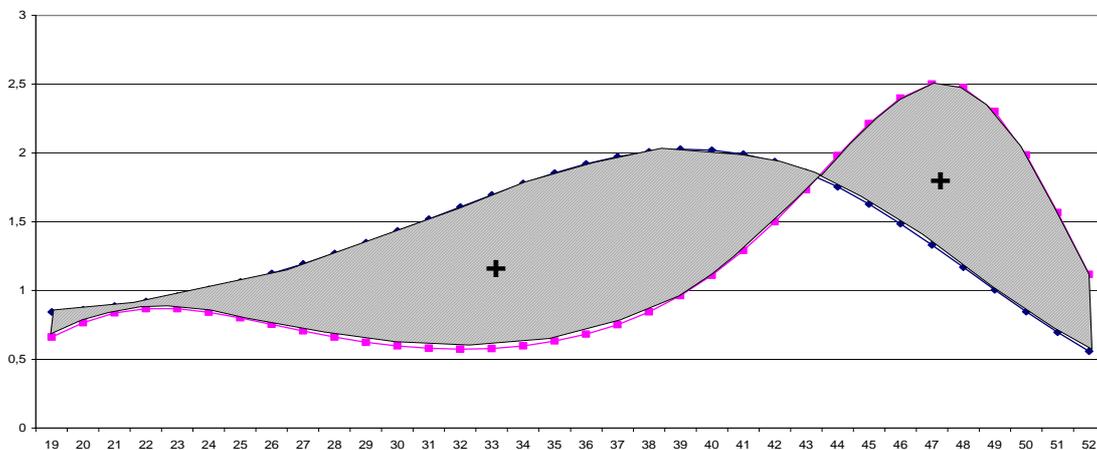
Plus les parcours salariaux de deux salariés se ressemblent, plus la distance entre leurs carrières sera petite. Cette distance est très discriminante puisqu'elle est nulle seulement si les deux salariés ont perçu exactement la même suite de salaires tout au long de leur carrière. Leurs carrières sont alors indistinguables. En revanche, elle est strictement positive dès que les salariés n'ont pas perçu le même salaire au moins une année. Cette distance capte donc plus d'éléments que les simples différences de niveau moyen de salaire sur toute la carrière (le salaire permanent). Elle permet par exemple de distinguer deux carrières qui auraient le même salaire permanent mais pas la même répartition de ce salaire entre les différents âges car ces deux carrières ne seraient alors pas à distance nulle l'une de l'autre.

La distance ainsi définie possède une propriété supplémentaire, celle de provenir d'un « produit scalaire » qui fait de l'espace des tendances polynomiales un espace « euclidien » possédant de bonnes propriétés. Etant donné deux carrières salariales f et g, le produit scalaire associé à la distance est simplement :

$$\langle f, g \rangle = E_e(f * g) = \frac{1}{T} \sum_1^T f_t \cdot g_t$$

où  $E_e$  désigne la moyenne empirique et T est la longueur de la carrière

Distance entre deux carrières : la contribution de chaque âge à la distance entre les deux carrières est toujours comptée positivement



## Les polynômes orthogonaux simplifient l'expression de la distance

Cette propriété importante permet de simplifier considérablement l'expression de la distance grâce à l'utilisation d'une base de polynômes orthogonaux. Chacune de ces fonctions élémentaires capture en effet une composante particulière, orthogonale à toutes les autres composantes, de la forme de la carrière. La décomposition des carrières individuelles sur une base de polynômes orthogonaux conduit ainsi à une expression plus simple de la distance.

Une base de polynômes  $(e_0, e_1, e_2, e_3, e_4)$  est dite orthonormale si les éléments de cette base sont de normés et deux à deux orthogonaux, propriétés qui se traduisent par les propriétés suivantes :

$$P_1 : \forall i, \langle e_i, e_i \rangle = 1$$

$$P_2 : \forall i \neq j, \langle e_i, e_j \rangle = 0$$

Lorsqu'on décompose (par les moindres carrés ordinaires) les deux tendances polynomiales dans une telle base, l'expression de la distance entre ces deux tendances prend précisément la forme « euclidienne » de la distance utilisée par les algorithmes de classification.

$$\left. \begin{array}{l} f = \sum_i C_i^f . e_i \\ g = \sum_i C_i^g . e_i \end{array} \right\} \Rightarrow dist(f, g)^2 = \|f - g\|^2 = \sum_i (C_i^g - C_i^f)^2$$

## Construction de la base de polynômes orthonormaux

Dans un espace euclidien la construction d'une base orthonormale est connue sous le nom de procédé d'orthonormalisation de Schmidt. Elle s'effectue de proche en proche en construisant un à un les vecteurs de la base souhaitée. Dans le cas particulier qui nous intéresse (celui d'une base orthonormale de polynômes), le procédé de Schmidt consiste à transformer la base non orthonormale formée des puissances du temps  $(1, t, t^2, t^3, t^4)$  en une nouvelle base orthonormale  $(e_0, e_1, e_2, e_3, e_4)$  possédant les propriétés voulues. Cette construction se fait par étapes successives.

Lors de la première étape, on se contente de poser  $e_0 = \frac{e}{\|e\|} = e$ , où  $e$  est le vecteur dont toutes les coordonnées sont égales à l'unité.

L'étape  $n$  consiste à construire  $e_{n-1}$ . Pour cela on effectue la régression suivante (par les moindres carrés ordinaires) :

$$t^{n-1} = (1, t, t^2, \dots, t^{n-2}) b_{n-1} + u_{n-1} \quad (1)$$

et on pose

$$e_{n-1} = \frac{\hat{u}_{n-1}}{\|u_{n-1}\|} = \frac{t^{n-1} - \hat{t}^{n-1}}{\|t^{n-1} - \hat{t}^{n-1}\|} \quad (2)$$

## La base de polynômes orthonormaux obtenue possède les propriétés annoncées

La forme du procédé garantit que les fonctions polynomiales de départ  $(1, t, t^2, t^3, t^4)$  et les fonctions orthonormalisées  $(1, t, t^2, t^3, t^4)$  engendrent le même espace :

$$\forall p, \langle 1, \dots, e^p \rangle = \langle 1, \dots, t^p \rangle$$

La relation (1) garantit de plus que  $\hat{u}_{n-1} \perp \langle 1, t, t^2, \dots, t^{n-2} \rangle = \langle 1, \hat{u}, \hat{u}^2, \dots, \hat{u}^{n-2} \rangle$ , ce qui implique la propriété  $P_2$ .

La relation (2) assure pour sa part que  $\|e_{n-1}\| = \frac{\|\hat{u}_{n-1}\|}{\|u_{n-1}\|} = 1$ , ce qui implique la propriété  $P_1$ .

### Possibilité de pondérer de manière différente les différentes années d'activité

La méthode s'appliquerait sans grand changement si on désirait prendre en compte les années de travail de manière différenciée ou lorsqu'on ne considère que certaines années (les 25 meilleures par exemple). En effet, il suffit de changer la distance considérée de la manière suivante :

$$dis_{tan ce_p}(f, g)^2 = E_e^p((f - g)^2) = \frac{1}{T} \sum_1^T p(t) \cdot (f_t - g_t)^2$$

où  $p$  est un poids vérifiant

$$\sum_1^T p(t) = 1$$

En pratique, le procédé de décomposition des carrières s'effectue quant à lui mutatis mutandis en remplaçant chaque carrière individuelle  $f_t$  par la carrière  $f'_t = f_t \sqrt{p(t)}$ . En effet, pour tout couple de fonctions  $(f, g)$ , le produit scalaire modifié par le poids  $p(t)$  est égal au produit scalaire habituel exprimé sur les fonctions modifiées par le poids  $p(t)$  :

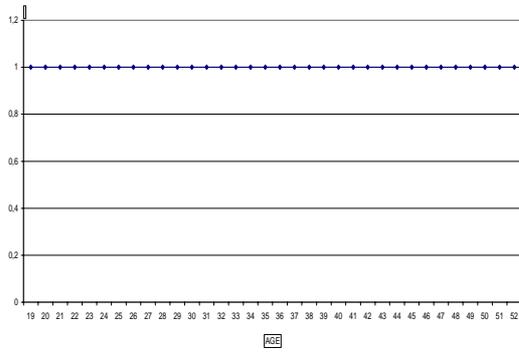
$$\langle f, g \rangle_p = E_e^p(f * g) = \frac{1}{T} \sum_1^T p(t) \cdot f_t \cdot g_t = \frac{1}{T} \sum_1^T (\sqrt{p(t)} \cdot f_t) \cdot (\sqrt{p(t)} \cdot g_t) = E_e(f' \cdot g') = \langle f', g' \rangle$$

Par exemple, pour tenir compte de la réforme Balladur de 1993 de la règle de fixation des retraites (calcul sur les 25 meilleures années), on ordonnerait la séquence des salaires perçus par chaque salarié dans un ordre croissant (et non plus dans un ordre chronologique) et on affecterait un poids  $p(t)=1/25$  pour les 25 dernières années et  $p(t)=0$  pour les autres.

## Annexe 3 : Les cinq polynômes orthogonaux obtenus par le procédé de Schmidt et utilisés pour décomposer les carrières individuelles

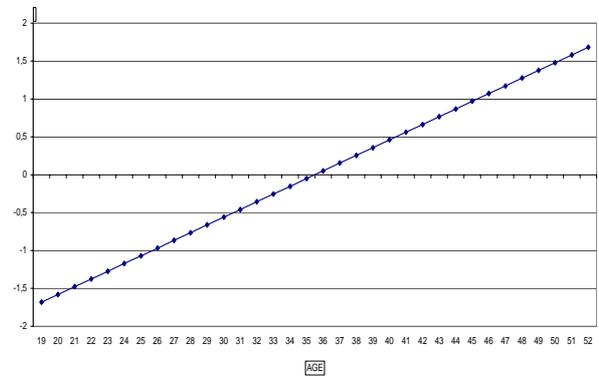
POLYNOME de degré 0

Polynômes formant la base de polynômes orthogonaux



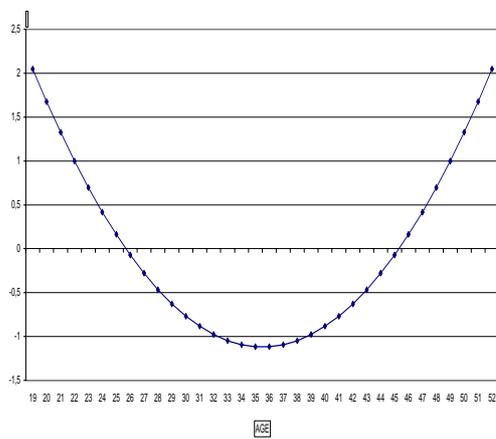
POLYNOME de degré 1

Polynômes formant la base de polynômes orthogonaux



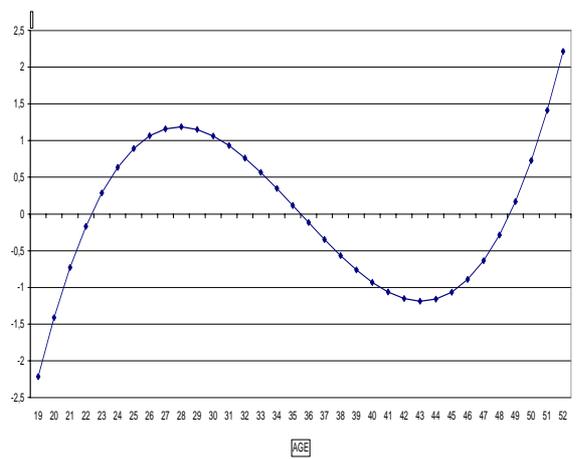
POLYNOME de degré 2

Polynômes formant la base de polynômes orthogonaux



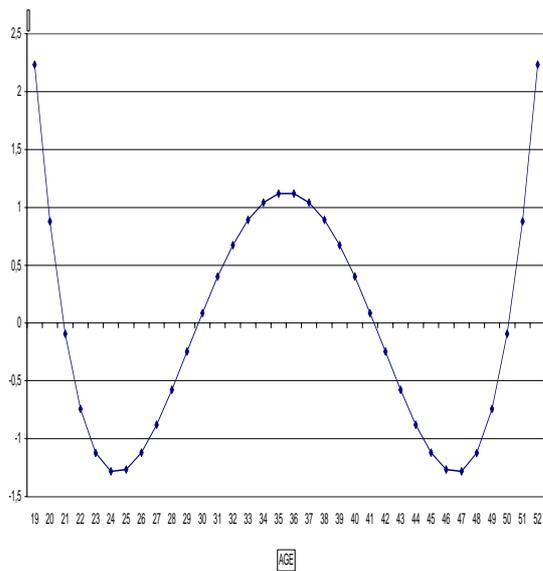
POLYNOME de degré 3

Polynômes formant la base de polynômes orthogonaux



POLYNOME de degré 4

Polynômes formant la base de polynômes orthogonaux



## Annexe 4 : Impact des caractéristiques individuelles sur les carrières salariales

Dans cette partie, on utilise une méthode complémentaire à celle des cas-types pour quantifier la diversité des carrières, méthode consistant à mesurer le lien existant entre la carrière salariale suivie par un salarié d'une part et ses caractéristiques individuelles observables d'autre part. Les résultats des parties précédentes ont déjà révélé quelques traits de ces corrélations : les différentes carrières-types obtenues dans la partie précédente ne sont pas également représentées parmi toutes les catégories de salariés ; de plus un événement ponctuel, comme le passage au statut de cadre, a des effets persistants sur la carrière, qui dépendent de surcroît du moment de la carrière où il intervient.

La méthode utilisée consiste à isoler l'effet propre de chaque caractéristique individuelle en mesurant l'écart de carrière entre un salarié possédant cette caractéristique et un salarié qui ne la possède pas, les autres caractéristiques étant par ailleurs fixées. Par rapport à une équation de gain classique, plusieurs facteurs supplémentaires sont pris en compte. Tout d'abord, la modélisation adoptée permet tout d'abord de mesurer l'effet des caractéristiques individuelles sur l'ensemble de la carrière et pas seulement à un instant donné. Ensuite elle permet de prendre éventuellement en compte l'hétérogénéité individuelle inobservée, qui fait que deux salariés ayant exactement les mêmes caractéristiques observables n'ont pas nécessairement le même salaire, ni la même progression de salaire, etc. Enfin on prend en compte l'existence éventuelle d'un biais de sélection. Celui-ci tient à la possible existence de caractéristiques individuelles inobservées influençant à la fois la probabilité d'être en activité sur le marché de l'emploi et le salaire lui-même, corrélation qui biaise l'estimation du gain engendré par telle ou telle caractéristique individuelle.

### L'impact sur la forme entière de la carrière

Pour prendre en compte l'effet des caractéristiques individuelles sur l'ensemble de la carrière salariale, on a systématiquement croisé toutes les caractéristiques individuelles avec les puissances de l'âge. L'idée qui sous-tend cette modélisation est que l'effet de l'âge sur la carrière est différent selon la catégorie de salariés considérée, ou, dit autrement, que les caractéristiques individuelles des salariés peuvent modifier le rendement de l'âge ou de l'expérience. Cette hypothèse est conforme aux constats déjà faits dans les parties précédentes pour ce qui concerne le sexe et l'accès au statut de cadre : il est par exemple notoire que l'expérience n'apporte pas le même bénéfice salarial pour un cadre que pour un ouvrier. Le jeu de variables utilisé peut s'écrire finalement de la manière suivante :

$$X' = (X, X \otimes age, X \otimes age^2, X \otimes age^3, X \otimes age^4)$$
$$X = (\text{sexe, condition d'emploi, cs, région})$$

On s'est limité aux puissances inférieures à cinq de l'âge, pour rester cohérent avec le choix du degré de l'approximation. Chacun des croisements de X avec une puissance de l'âge mesure l'impact de X sur une des composantes géométriques de la carrière : niveau, pente, et évolutions d'ordre supérieur (retournements, etc.).

### Prendre en compte l'hétérogénéité inobservée ?

Pour mesurer l'impact des caractéristiques individuelles observables sur la carrière salariale, on peut mesurer l'écart de salaire moyen entre ceux qui possèdent cette caractéristique et ceux qui ne la possèdent pas, les autres caractéristiques observables étant égales par ailleurs. C'est la perspective - essentiellement descriptive - dans laquelle on se place ici. Il est également possible de prendre en compte dans le modèle l'hétérogénéité individuelle inobservée. L'effet estimé correspondrait dans ce cas à l'écart de salaire induit *pour un même individu* entre les situations où il possède la caractéristique et celles où il ne la possède pas. Le « toutes choses égales par ailleurs » s'étend dans ce cas à la personne même du salarié.

Bien que dans la perspective descriptive dans laquelle on se place, on s'intéresse particulièrement à l'estimateur "between", il n'est toutefois pas inintéressant de calculer l'estimateur "within", car celui-ci permet classiquement de tester l'endogénéité de certaines variables et apporte ainsi une information complémentaire sur le comportement des agents.

Dans l'estimation « within », l'hétérogénéité individuelle inobservée est classiquement modélisée à l'aide d'effets fixes individuels. Ceux-ci sont sensés représenter la part du salaire explicable seulement par des caractéristiques inobservables propres à l'individu considéré. La partie du modèle comportant ces effets fixes constitue en quelque sorte « l'équation personnelle » de chaque salarié. Dans la plupart des cas, on suppose que la personnalité propre du salarié n'a d'influence que sur la moyenne de son salaire mesurée durant sa carrière (son salaire permanent). Cette hypothèse pratique tient au faible nombre d'années d'observation dont on dispose pour chaque salarié et qui ne permettent pas de modéliser de manière très élaborée l'effet propre de l'individualité du salarié sur sa carrière (voir encadré 2). Grâce à la multiplicité d'observations dont on dispose ici pour chaque individu, on peut modéliser de manière plus complète l'équation de manière à ce que soit pris en compte l'influence des caractéristiques inobservées sur d'autres aspects de son profil de carrière que le seul salaire permanent. L'effet de carrière propre à l'individu consistera ici en un polynôme de degré 4. Les effets fixes individuels sont croisés avec les puissances de l'âge :

$$I' = (b_i, b_i * age, b_i * age^2, b_i * age^3, b_i * age^4)_{i=1..N}$$

**Encadré 2 : Méthode pour prendre en compte des caractéristiques individuelles inobservées**

L'impact d'une caractéristique donnée peut être mesuré de différentes manières, selon la nature du phénomène que l'on désire mesurer et les autres paramètres que l'on désire prendre en compte. Prenons le cas du statut de cadre. On veut mesurer « toutes choses égales par ailleurs » le différentiel de salaire induit par le fait d'être cadre. Une approche descriptive qui s'en tient aux caractéristiques observables consistera à comparer les salaires des cadres et ceux des non-cadres, à autres caractéristiques observables - telles que le sexe - fixées. On peut vouloir en plus tenir compte de l'hétérogénéité individuelle inobservée existant entre les salariés. Le « toutes choses égales par ailleurs » s'étend alors à la personne même du salarié et il convient alors d'évaluer le changement de salaire induit *pour un même salarié* entre des situations où il est cadre et des situations où il est non-cadre.

Le premier type d'estimation est dit « inter-individuelle ». L'estimateur « between » correspondant est descriptif, il est adapté pour mesurer des écarts de salaire moyens entre différentes catégories de salariés. Si  $I$  désigne un ensemble de variables modélisant les caractéristiques individuelles inobservées, et  $P_I$  le projecteur sur l'espace engendré par les variables  $I$ , l'estimateur between  $b_b$  correspond à la régression suivante :

$$P_I Y_{i,t} = P_I X_{i,t} \cdot b_b + v_{i,t} \tag{1}$$

L'estimateur « within »  $b_w$  réalise quant à lui l'estimation « intra-individuelle » incluant dans le modèle les variables modélisant l'hétérogénéité individuelle. Il mesure l'écart de salaire induit par le fait d'être cadre *pour un même individu*. L'équation estimée s'écrit alors en écart à l'« équation personnelle » du salarié :

$$(Id - P_I) Y_{i,t} = (Id - P_I) X_{i,t} \cdot b_w + u_{i,t} \tag{2}$$

En pratique, on fait la synthèse de ces deux approches en estimant la régression (3) qui donne les deux estimateurs cherchés. L'estimation conjointe des deux estimateurs permet de plus de tester leur égalité ou leur différence. Celle-ci est d'autant plus significative que l'hétérogénéité individuelle  $v_{i,t}$  du modèle « between » est corrélée aux caractéristiques observables  $X_{i,t}$  dont on cherche à mesurer l'effet.

$$Y_{i,t} = X_{i,t} \cdot b_w + P_I \cdot X_{i,t} \cdot (b_b - b_w) + w_{i,t} \tag{3}$$

## Le traitement des effets de sélection

La sélection endogène tient au fait qu'un salarié peut, en fonction du salaire qu'il anticipe, choisir de ne pas participer au marché de l'emploi. Or certaines caractéristiques non directement observables de l'individu influencent à la fois son choix de participer et le niveau de son salaire et peuvent de ce fait être vues comme des variables omises. Il en résulte un biais dans l'estimation des coefficients de la fonction de gain, dans la mesure où ces variables omises sont corrélées avec le résidu de l'équation de gain. Les rendements estimés des caractéristiques individuelles incluent alors une partie du gain tenant au simple fait d'être présent sur le marché de l'emploi.

Pour corriger ce biais, la procédure habituellement utilisée en coupe, due à Heckman, consiste à inclure parmi les régresseurs une fonction du résidu d'une équation de participation. Guillotin et Sevestre (1994) notent que la procédure est lourde à mettre en œuvre sur données de panel. Elle a aussi l'inconvénient de modéliser de manière unidimensionnelle le processus de sélection qui peut en fait résulter de la superposition de phénomènes hétérogènes : jeunes effectuant leur service national, périodes de chômage, congé maternité, sortie des chefs d'entreprises salariés vers le champ des indépendants, etc). La méthode de Nijman Verbeeck (1992) que l'on utilise ici consiste à ajouter au modèle des variables liées au statut de participation du salarié. Cette méthode ne suppose pas de processus de sélection unique et approche de manière satisfaisante la correction mise en jeu par la méthode d'Heckman. Les régresseurs supplémentaires que l'on a introduits sont les suivants :

NPRES : nombre d'années de présence

PREC1 : 1 si présent l'année n-1, 0 sinon

CC : 1 si sortie définitive avant la dernière année observée (l'année 2000 ici) et après une carrière ininterrompue, 0 sinon

## Effets des principales caractéristiques sur la forme des carrières salariales

La simulation de l'impact des différentes caractéristiques individuelles sur l'ensemble de la carrière est présentée dans les graphiques 21 à 23. On a d'autre part estimé cet impact de manière plus synthétique, sur la partie de carrière située entre 30 et 50 ans et en séparant l'effet de chaque caractéristique individuelle sur le niveau moyen de salaire (salaire permanent) de son effet sur la croissance du salaire sur cette période (Tableaux 24 à 26) Chacune des estimations a été faite séparément pour l'ensemble des salariés, pour les salariés des deux sexes pris séparément. Comme pour les méthodes classiques de régression, il est nécessaire de fixer une carrière de référence. Il s'agit en l'occurrence de celle d'un employé de sexe masculin, exerçant son activité à temps plein en province.

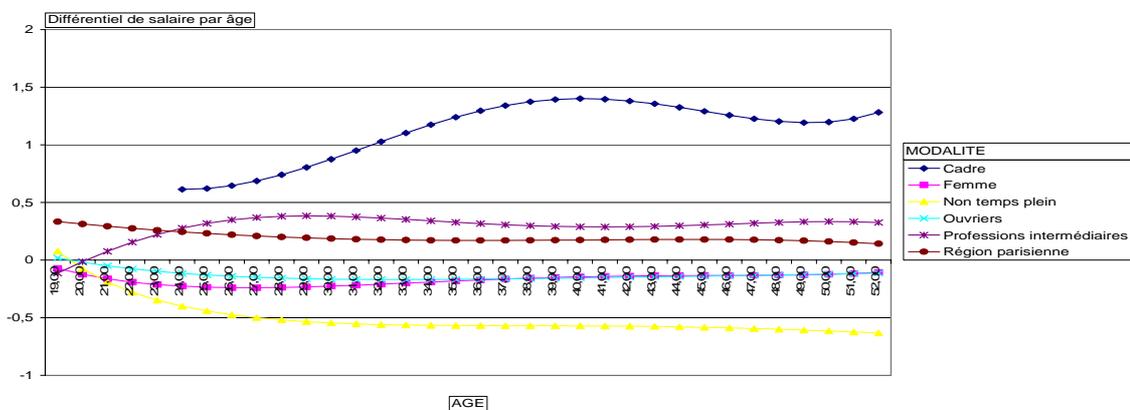
Limitons-nous d'abord, dans une perspective descriptive à l'estimation interindividuelle (Between), qui met l'accent sur les différences observables entre les salariés. Les cadres font la différence en première partie de carrière. L'écart entre leur salaire et celui correspondant à la carrière de référence croît jusqu'à 40 ans, puis se stabilise, voire diminue légèrement (Graphique 21). Entre 30 et 50 ans, le salaire des cadres est plus élevé de 245 % que la carrière de référence et il croît de 38 points de plus que celle-ci (Tableau 24). L'avantage pour les professions intermédiaires se stabilise plus tôt, entre 28 et 30 ans. L'écart de salaire entre les hommes et les femmes commence par croître en début de carrière, puis diminue ensuite. Au total, évaluée entre 30 et 50 ans (Tableaux 24 à 26), les carrières des femmes sont « toutes choses observables égales par ailleurs » plus basses que celles des hommes (-15% sur le salaire permanent) mais aussi plus croissantes (+10 points de croissance du salaire entre 30 et 50 ans). Quant aux conditions d'emploi particulières, elles entraînent une perte de salaire qui s'amplifient avec l'âge. Entre 30 et 50 ans, les salariés occupant une condition d'emploi particulière (autre que temps plein) gagnent en moyenne 44% de moins et leur salaire croît de 7 points de moins que la carrière de référence. Le fait d'exercer sa profession en région parisienne procure en revanche à lui seul un bénéfice salarial de 19%. En revanche, les carrières parisiennes sont légèrement moins croissantes que la référence (-2,5 points).

Les principales caractéristiques individuelles engendrent des écarts de carrière qui ne sont pas les mêmes pour les hommes et pour les femmes. Ainsi, le fait d'occuper un poste de cadre ou une profession intermédiaire est moins rentable pour les femmes. Pour un homme, le fait d'être cadre entraîne une plus-value de 65% à 25 ans, avantage porté à 140% à 50 ans (Graphique 23). Pour les femmes, non seulement cet avantage est moindre, mais il croît aussi moins avec l'âge. Au total, évalué entre 30 et 50 ans, le fait d'être cadre rapporte un différentiel de salaire permanent de 317% pour les hommes contre 198% pour les femmes (Tableaux 25 et 26). Sur cette période, la carrière des cadres hommes croît de 65% de plus que la carrière de référence, alors qu'elle croît de 1,8% de moins pour les femmes. Ce dernier point est toutefois à rapprocher du fait que les carrières des femmes (et en particulier la carrière de référence du modèle) sont en moyenne plus croissantes que celle des hommes. Le fait d'exercer sa profession dans une autre condition d'emploi qu'à temps plein est moins pénalisant pour les femmes que pour les hommes, surtout du point de vue de la croissance du salaire. Le statut d'ouvrier est en revanche plus pénalisant pour les femmes que pour les hommes, tant du point de vue du salaire permanent que de sa croissance. Le fait d'exercer en région parisienne avantage les femmes par rapport aux hommes du point de vue du salaire permanent, mais les désavantage du point de vue de la croissance du salaire.

La prise en compte de l'hétérogénéité inobservée modifie substantiellement ces estimations. Les écarts de salaire mesurés correspondent en effet à l'effet induit par une caractéristique donnée mesuré *sur un même individu*. La plupart des estimateurs sont plus faibles en valeur absolue, ce qui incite à penser que les caractéristiques inobservées jouent un rôle important sur la forme des carrières. Ainsi, l'avantage induit par le fait d'être cadre sur le salaire permanent entre 30 et 50 ans (estimateur « within » du Tableau 24) n'est « que » de 11,4%. Cela signifie que les salariés promus au statut de cadre ne sont pas tirés au hasard, mais parmi ceux qui avaient déjà des salaires élevés pour leur catégorie d'appartenance, de sorte que l'accès au statut de cadre n'a en lui-même entraîné qu'une faible augmentation de salaire.

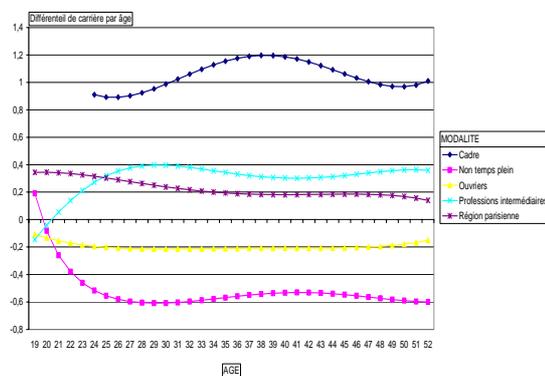
TYPE|Between

Différentiel de carrière induit "toutes choses égales par ailleurs" par certaines caractéristiques individuelles



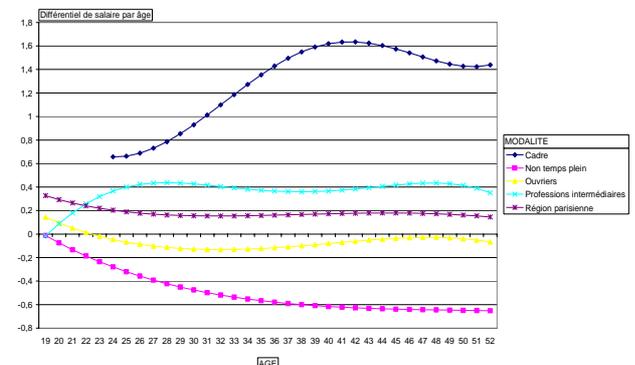
TYPE|Between

Différentiel de carrière induit "toutes choses égales par ailleurs" par certaines caractéristiques individuelles



AGE

TYPE|Between



AGE

**Graphiques 21 à 23** : Ecart de carrière engendré par différentes caractéristiques individuelles, par sexe.

## Ensemble

	Données		TYPE		Salaire permanent	
	Pente					
MODALITE	Between	Within	Between	Within	Between	Within
Cadre		0,381	0,012		2,449	0,114
Femme		0,108			-0,147	
Non temps plein		-0,068	-0,148		-0,438	-0,303
Ouvriers		0,044	-0,026		-0,140	0,010
Professions intermédiaires		-0,046	-0,021		0,379	0,055
Région parisienne		-0,025	0,056		0,191	0,035

## Femmes

	Données		TYPE		Salaire permanent	
	Pente					
MODALITE	Between	Within	Between	Within	Between	Within
Cadre		-0,018	0,051		1,981	0,083
Non temps plein		0,018	-0,086		-0,430	-0,282
Ouvriers		0,036	0,006		-0,186	-0,011
Professions intermédiaires		-0,033	-0,002		0,402	0,032
Région parisienne		-0,068	-0,015		0,211	0,026

## Hommes

	Données		TYPE		Salaire permanent	
	Pente					
MODALITE	Between	Within	Between	Within	Between	Within
Cadre		0,645	-0,009		3,171	0,144
Non temps plein		-0,159	-0,179		-0,449	-0,318
Ouvriers		0,093	-0,042		-0,078	0,033
Professions intermédiaires		-0,012	-0,036		0,486	0,082
Région parisienne		0,007	0,080		0,182	0,039

**Tableaux 24 à 26** : Ecart de salaire permanent et de croissance du salaire engendré entre 30 et 50 ans par différentes caractéristiques individuelles, par sexe. Estimateurs inter-individuel (Between) et intra-individuel (Within)