JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

MICHEL GLAUDE

L'utilisation des modèles d'analyse de variance dans l'étude des salaires et des carrières

Journal de la société statistique de Paris, tome 130, nº 2 (1989), p. 63-79

<http://www.numdam.org/item?id=JSFS_1989__130_2_63_0>

© Société de statistique de Paris, 1989, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (http://publications-sfds.math.cnrs.fr/index.php/J-SFdS) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

II

COMMUNICATION

L'UTILISATION DES MODÈLES D'ANALYSE DE VARIANCE DANS L'ÉTUDE DES SALAIRES ET DES CARRIÈRES [1]

MICHEL GLAUDE
administrateur de l'INSEE,
chef de la division « Revenus »

Cet article propose deux exemples d'utilisation de modèles d'analyse de variance, concernant les salaires de l'ensemble des salariés français en 1987 et les salaires des anciens élèves diplômés des écoles d'ingénieurs. On y insiste plus particulièrement sur les effets de structure de population et on y présente sous forme de graphiques le passage des disparités apparentes entre catégories de salariés telles qu'elles résultent de la simple tabulation, aux disparités nettes ou résiduelles telles qu'elles ressortent du modèle d'analyse de variance. Enfin, l'étude des disparités liées à l'expérience professionnelle fournit une approche des effets de carrière hors mouvements conjoncturels.

Variance analysis applied to wages and carreer studies.

Two examples of variance analysis are presented. The first one is based upon wages data taken from the french labor force survey of 1987 and the second one concern former students graduate from the french specific schools of ingeneering.

Effects of population structure are commented and lead from gross discrepancies between groups of population, as yield by a simple tabulation, to net discrepancies yield by the variance analysis.

Differences along the live cycle introduce net career effects which are discussed.

Si la théorie des modèles multivariés est déjà ancienne, leur mise en pratique et leur utilisation par le statisticien de base l'est déjà beaucoup moins. Dans ce cas comme dans nombre d'autres, il s'avère que l'utilisation effective d'une nouvelle méthode est largement dépendante de la disponibilité de programmes de calcul standard facilement accessibles au statisticien.

L'apport des analyses multivariées est pourtant considérable en matière de statistique descriptive pour peu que la variable d'intérêt dépende de nombreux critères plus ou moins liés entre eux. C'est manifestement le cas du salaire où interviennent des facteurs tenant à l'individu comme à l'entreprise où à son environnement [2]. Deux exemples, l'un concernant l'ensemble des salariés et l'autre les anciens élèves diplômés des écoles d'ingénieurs, nous permettront de mettre en évidence l'intérêt de cette méthode et plus particulièrement son application à l'étude des effets de carrière.

I. L'ÉVENTAIL DES SALAIRES EN 1987

En mars 1987, les salariés à temps complet déclaraient à l'enquête sur l'emploi de l'INSEE un salaire net moyen de l'ordre de 6 800 francs par mois. Les 10 % les moins bien payés recevaient moins

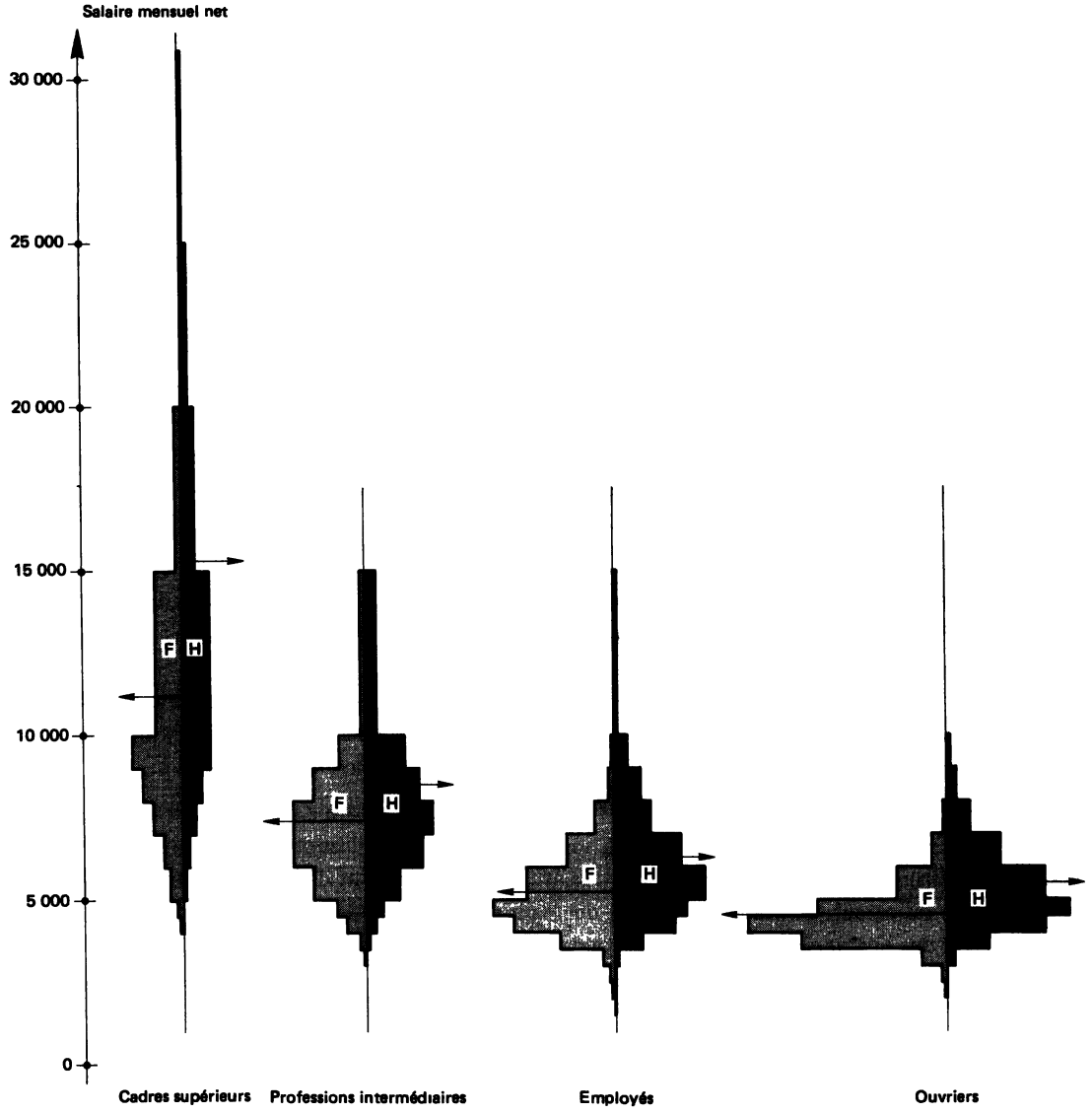
1. Ce texte reprend une grande partie de la conférence présentée le 25 mai 1988 sous l'égide des Sociétés de statistique de Paris et de France et de l'ASTEC et intitulée : « Salaire et profil de carrière : le cas des jeunes et des ingénieurs. »

2. C'est bien aussi le cas de variables qualitatives comme par exemple le fait d'être au chômage, et la diffusion récente des modèles qualitatifs multivariés procède du même mouvement que celle de l'analyse de variance commentée ici.

Journal de la Société de statistique de Paris, tome 130, N° 2, 1989.

Figure 1

Pyramide des salaires en 1987 par sexe et catégorie sous-professionnelle*



* Source : Enquête sur l'emploi de mars 1987.

de 4 100 francs par mois et environ 7 % se situaient en dessous du SMIC mensuel net calculé sur la base de 39 heures, soit 3 900 francs par mois.

A l'autre extrémité, les 10 % de salariés les mieux payés gagnaient plus de 10 200 francs net par mois, 1 % déclarant gagner plus de 24 000 francs mensuels. Le salaire médian, partageant la population des salariés en deux, se situait aux environs de 5 800 francs net par mois.

(Pour plus de précisions quant à la fiabilité de ces données collectées par interview par rapport aux données issues des déclarations des employeurs, on se reportera à l'encadré p. 170 de [1]).

Les salaires masculins sont supérieurs aux salaires féminins de 24 % en moyenne (7 340 francs par mois contre 5 900 francs). Ils sont également plus dispersés, l'éventail des rémunérations masculines est plus ouvert que celui des rémunérations féminines.

L'étude par grande catégorie socioprofessionnelle et par sexe semble montrer que plus le salaire moyen d'une catégorie est élevé plus la diversité des situations est importante. Ainsi les salaires ouvriers sont faibles (5 500 francs pour la moyenne des hommes et 4 500 francs pour les femmes) alors que les cadres, surtout masculins, ont des salaires élevés mais aussi beaucoup plus dispersés : 4 % des hommes cadres supérieurs déclarent gagner plus de 30 000 francs par mois (voir figure 1).

1. *Les effets de structure*

Cette description des disparités de salaires pourrait être poursuivie par l'étude détaillée de l'influence de chaque criètre disponible à l'enquête sur l'emploi : région de résidence, secteur d'activité et taille de l'établissement, nationalité du salarié... Outre la lassitude qui s'ensuivrait, cette démarche n'est pas à l'abri de répétitions qui tiennent à ce qu'on appelle les effets de structure, et qui peuvent même aller jusqu'à fausser l'analyse.

Prenons le cas des disparités régionales et supposons par exemple que les entreprises de la région parisienne paient leurs salariés 20 % de plus que les entreprises de province. En admettant que les cadres gagnent deux fois plus que les autres employés et que la proportion de cadres soit de 20 % en région parisienne contre 10 % en province. On obtient alors un rapport du salaire moyen de la région parisienne à la province de $1{,}31 = 1{,}2\ [0{,}8 + 0{,}2 \times 2]/[0{,}9 + 0{,}1 \times 2]$.

Ainsi l'effet de structure, une présence plus forte des cadres en région parisienne, aurait produit une disparité régionale apparente (31 %) plus importante que la disparité régionale effective (20 %) et ce qu'un commentaire brutal attriburait à la région serait pour partie dû aux effets de la qualification. L'analyse de variance utilisée dans cette étude vise à éviter les pièges des disparités apparentes pour essayer de restituer des disparités réelles ou effets spécifiques de chaque variable. Mais comme toute méthode synthétique, sa puissance repose sur des conventions et la production des disparités nettes dépend du modèle utilisé et de la liste des variables explicatives retenues.

2. *Un modèle multiplicatif comportant de nombreuses variables explicatives*

La spécification retenue pour l'analyse de variance est multiplicative (ou additive en logarithme) et ne comporte pas, dans un premier temps, d'interactions entre variables explicatives. C'est faire l'hypothèse que le salaire d'un individu donné, c'est-à-dire présentant un ensemble spécifique de caractéristiques, peut se décomposer en une suite (dans notre cas multiplicative) d'effets tenant à chaque caractéristique et identiques pour tous les individus présentant cette caractéristique. Par exemple, cela revient à supposer que le fait de travailler en région parisienne accroît systématiquement le salaire de x % que l'on soit manœuvre ou cadre, que l'on travaille dans un petit ou grand établissement... (voir annexe).

Cette hypothèse d'absence d'interactivité des critères qui peut sembler assez forte est en fait bien vérifiée dans la réalité, pour l'ensemble des critères d'analyse excepté pour le sexe du salarié. C'est parce que les effets du secteur, de la qualification, du diplôme... ne sont pas identiques pour les hommes et pour les femmes que nous avons effectué deux analyses séparées pour les salaires masculins et féminins.

Treize variables explicatives comportant 182 modalités ont été sélectionnées (Tableau 1). Elles renvoient à différentes théories du marché du travail selon qu'elles concernent plutôt le salarié, côté demande, l'emploi, côté offre, ou qu'elles caractérisent le marché du travail (externe et interne). Mais

TABLEAU 1
Liste des variables retenues dans les analyses de disparités de salaire

A — *Tous salariés (enquête sur l'emploi mars 1987)*	
Variables retenues	Nombre de modalités
Caractéristiques du salarié	
Sexe	2
Nationalité	6
Diplôme le plus élevé obtenu	14
Type de ménage	7
Expérience professionnelle totale	8
Caractéristiques de l'emploi	
Durée du travail	11
Qualification, catégorie socioprofessionnelle	23
Fonction (pour les cadres)	15
Secteur d'activité de l'établissement	39
Taille de l'établissement	11
Type d'entreprise du secteur public et statut du salarié du privé	10
Caractéristique du marché du travail (externe et interne)	
Région de résidence	21
Catégorie de commune de résidence	7
Ancienneté dans l'entreprise	8

B — *Anciens élèves masculins diplômés d'une école d'ingénieur* (enquête FASFID 1987)	
Variables retenues	Nombre de modalités
Caractéristiques du salarié	
École de formation	33
Type de ménage	8
Expérience professionnelle totale	7
Caractéristiques de l'emploi	
Position hiérarchique	13
Activité dominante	10
Secteur d'activité de l'entreprise	22
Taille de l'entreprise	5
Statut de l'entreprise	6
Caractéristiques du marché du travail	
Région d'emploi	4
Ancienneté dans l'entreprise	8

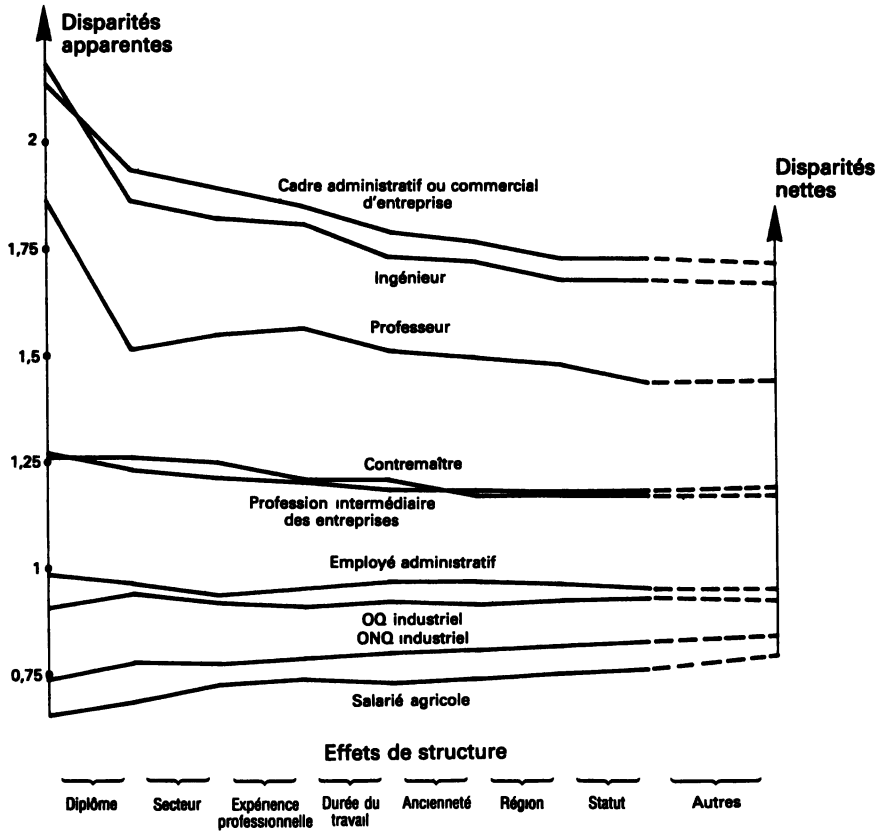
de nombreuses variables peuvent être porteuses de plusieurs interprétations : ainsi le diplôme peut représenter tout autant un niveau de formation (théorie du capital humain) qu'il peut signaler pour l'employeur un « bon élément » qui, ayant bien effectué son parcours scolaire, devrait faire de même dans l'entreprise (théorie du filtre).

3. *Une mesure du pouvoir explicatif spécifique de chaque critère*

Plus des deux tiers de la variabilité des salaires masculins (68 %) et un peu moins pour les salaires féminins (59 %) sont expliqués par le modèle d'analyse de la variance. La catégorie socioprofessionnelle du salarié est la variable la plus explicative des disparités de salaire. Plusieurs indicateurs peuvent être

Figure 2

Des disparités apparentes aux disparités nettes
Le cas de la catégorie socio-professionnelle



Lire ainsi : L'effet de structure du diplôme fait passer la disparité apparente des ingénieurs de 2,19 à 1,86, puis l'effet du secteur fait passer la disparité ingénieur de 1,86 à 1,82... Une fois tous les effets de structure décomptés, on obtient la disparité nette égale à 1,67.

utilisés pour mesurer l'apport de chaque variable à l'explication. Le critère que nous avons retenu utilisés pour mesurer l'apport de chaque variable à l'explication. Le critère que nous avons retenu mesure la perte de variance expliquée que cause la suppression de cette variable par rapport au modèle complet (voir annexe).

Ainsi, enlever le critère catégorie socioprofessionnelle du modèle ferait baisser le pourcentage de variance expliquée de 8 points pour les hommes et de 10 points pour les femmes. A un deuxième niveau d'importance du pouvoir explicatif spécifique des critères interviennent le diplôme pour les hommes et le diplôme et l'ancienneté dans l'entreprise pour les femmes.

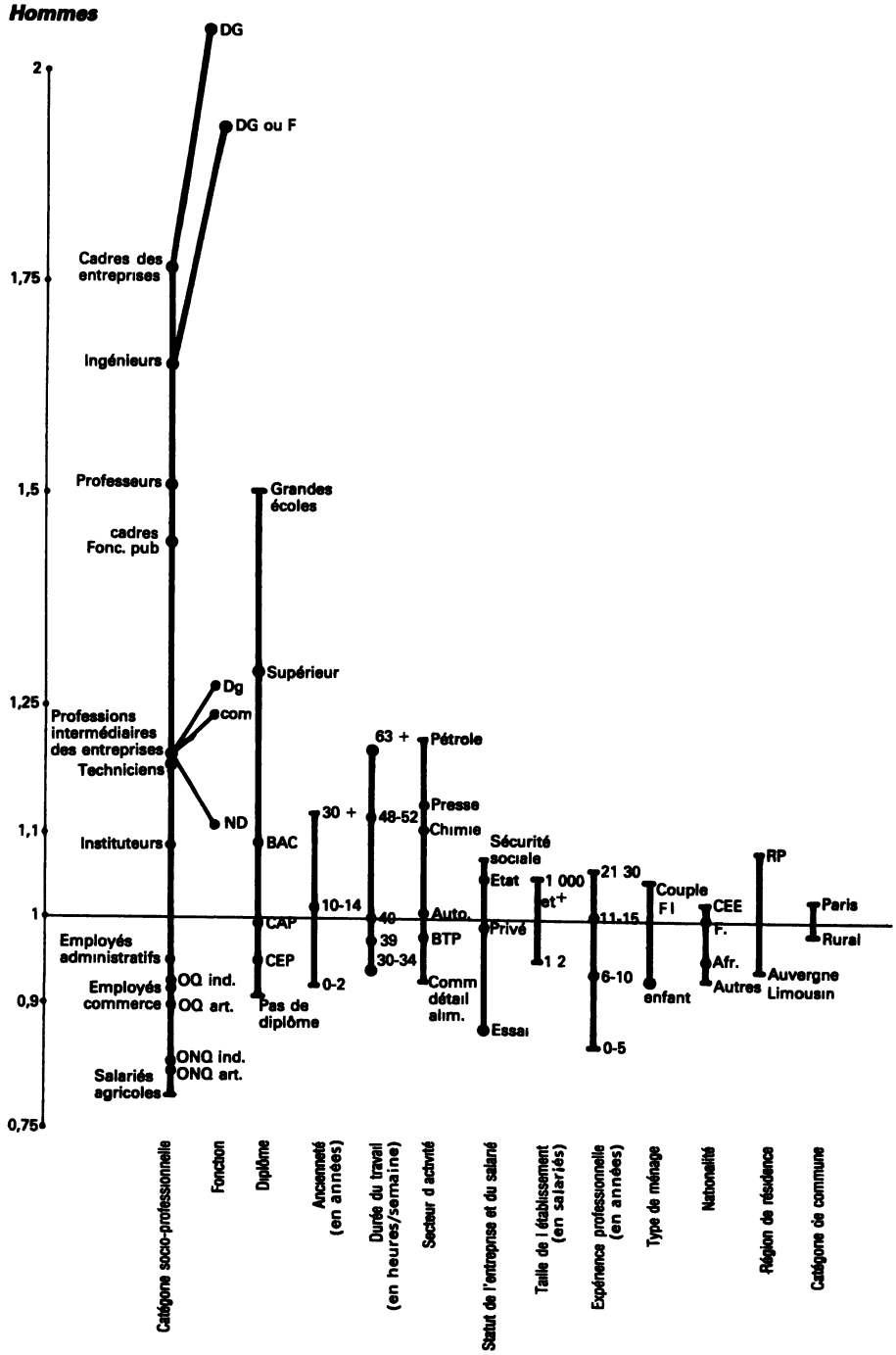
4. *Des disparités apparentes aux disparités nettes*

Les disparités apparentes de salaire entre groupes socioprofessionnels sont très importantes. Ainsi l'ingénieur est-il au niveau 221 et le salarié agricole au niveau 66 si 100 désigne le niveau moyen.

Figure 3

Disparités nettes de salaire 1987

Hommes



Chez les femmes l'éventail est aussi ouvert : près de 200 pour la femme cadre administratif d'entreprise contre 63 pour la femme personnel de service.

Les effets de structure réduisent ces écarts d'un bon tiers. On enlève en somme des disparités apparentes par catégorie socioprofessionnelle ce qui peut être expliqué par d'autres variables. Ainsi une partie du salaire des cadres est constitué par la rémunération du diplôme, de l'expérience professionnelle... Une fois décompté ces effets on obtient les disparités nettes que l'on pourrait aussi qualifier de pures ou résiduelles. Ainsi chez les hommes l'effet spécifique attribué au fait d'être ingénieur ne compte-t-il plus que pour 162 contre 220 pour la disparité apparente est l'effet salarié agricole passe de 66 à 79 après correction des effets de structure.

Compte tenu des liaisons entre catégorie socioprofessionnelle et diplôme, c'est l'effet de structure dû au diplôme qui est le plus notable. Ainsi pour les hommes l'effet du diplôme réduit de 24 % la disparité apparente des professeurs et de 19 % de celle des ingénieurs... Les différents effets de structure sont détaillés pour les salaires masculins par catégorie sociale (figure 2). Ils permettent de passer des disparités apparentes aux disparités nettes et peuvent être positifs ou négatifs. Leur ordre n'a pas d'importance, on a seulement visualisé les plus importants en commençant par eux (diplôme, secteur, expérience totale...).

5. *Des résultats du modèle à la prévision*

Sur la figure 3 ont été visualisés les effets spécifiques de chaque critère d'analyse correspondant aux disparités nettes estimées par le modèle pour les salariés masculins (les données correspondantes figurent au tableau n° 2). On y retrouve l'importance du critère qualification de l'emploi occupé qui en plus a été amélioré grâce à la prise en compte de la fonction exercée pour les cadres. Ainsi occuper un emploi de Directeur Général ou Financier quand on est cadre dans une entreprise apporte toutes choses égales par ailleurs un bonus qui double le salaire moyen. Si en plus ce cadre est diplômé d'une grande école il faut y ajouter 50 %, etc. pour chaque critère d'analyse.

Inversement, en partant des valeurs des disparités nettes estimées pour chaque critère on peut reconstituer le salaire estimé d'un individu défini par la série de ses caractéristiques individuelles. Chacun peut ainsi calculer quel aurait été son salaire mensuel en 1987, et le comparer à sa feuille de paie...

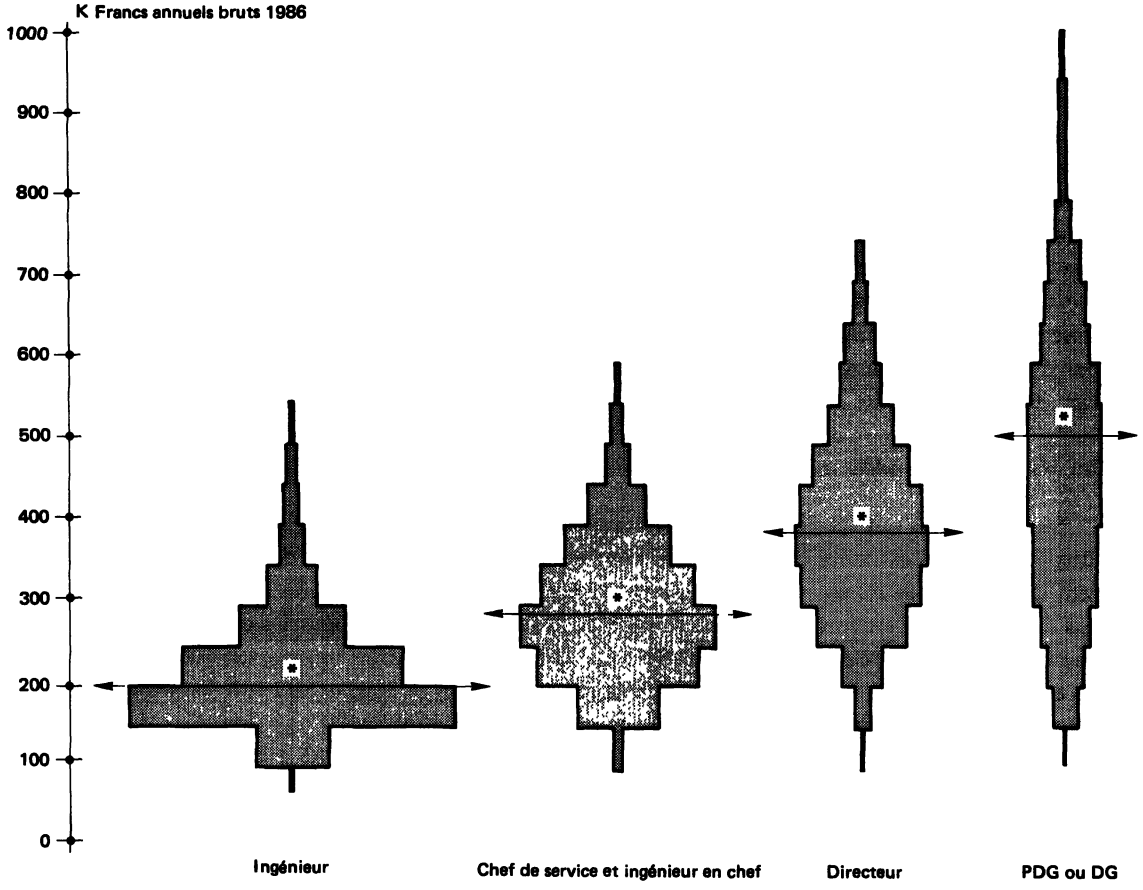
Notons que si la qualité de l'estimation du modèle d'analyse de variance est très bonne (2/3 de variance expliquée pour quelque 25 000 salariés masculins) le calcul d'un intervalle de confiance autour du salaire estimé laisse subsister une bonne marge d'incertitude due à l'ensemble des causes individuelles non prises en compte dans le modèle.

Si S est le salaire estimé le salaire réel aura :
2 chances sur 3 d'être compris dans l'intervalle [−20 %, +27 %]
4 chances sur 5 d'être compris dans l'intervalle [−27 %, +36 %]
et si on ne se laisse qu'une marge d'erreur de 5 % il faut étendre l'intervalle de confiance à [−39 %, +63 %] (Moins on veut de chances de se tromper plus la marge d'incertitude s'accroît).

II. Le salaire des anciens élèves diplômés d'une école d'ingénieurs

Pour mieux appréhender le haut de la distribution des salaires et les facteurs les plus contributifs des disparités salariales, nous avons analysé les données collectées en 1987 par la FASFID (Fédération des Associations et Sociétés Françaises d'Ingénieurs Diplômés) sur les niveaux de rémunération des anciens élèves des écoles d'ingénieurs pour l'année 1986. Plus de 20 000 ingénieurs salariés en activité ont répondu à cette enquête postale effectuée tous les 3 ou 4 ans depuis 1958. Compte tenu du faible nombre de femmes anciennes élèves, on s'est limité à l'analyse des rémunérations masculines qui sont

Figure 4
Pyramide des salaires des ingénieurs diplômés en 1986 selon la position hiérarchique*



* Source : Enquête FASFID 1987.
⟷ : médiane.

ici mesurées brutes de cotisations sociales à la charge du salarié et qui comprennent également les éventuels gains pour activité secondaire « liée à la qualité d'ingénieur ».

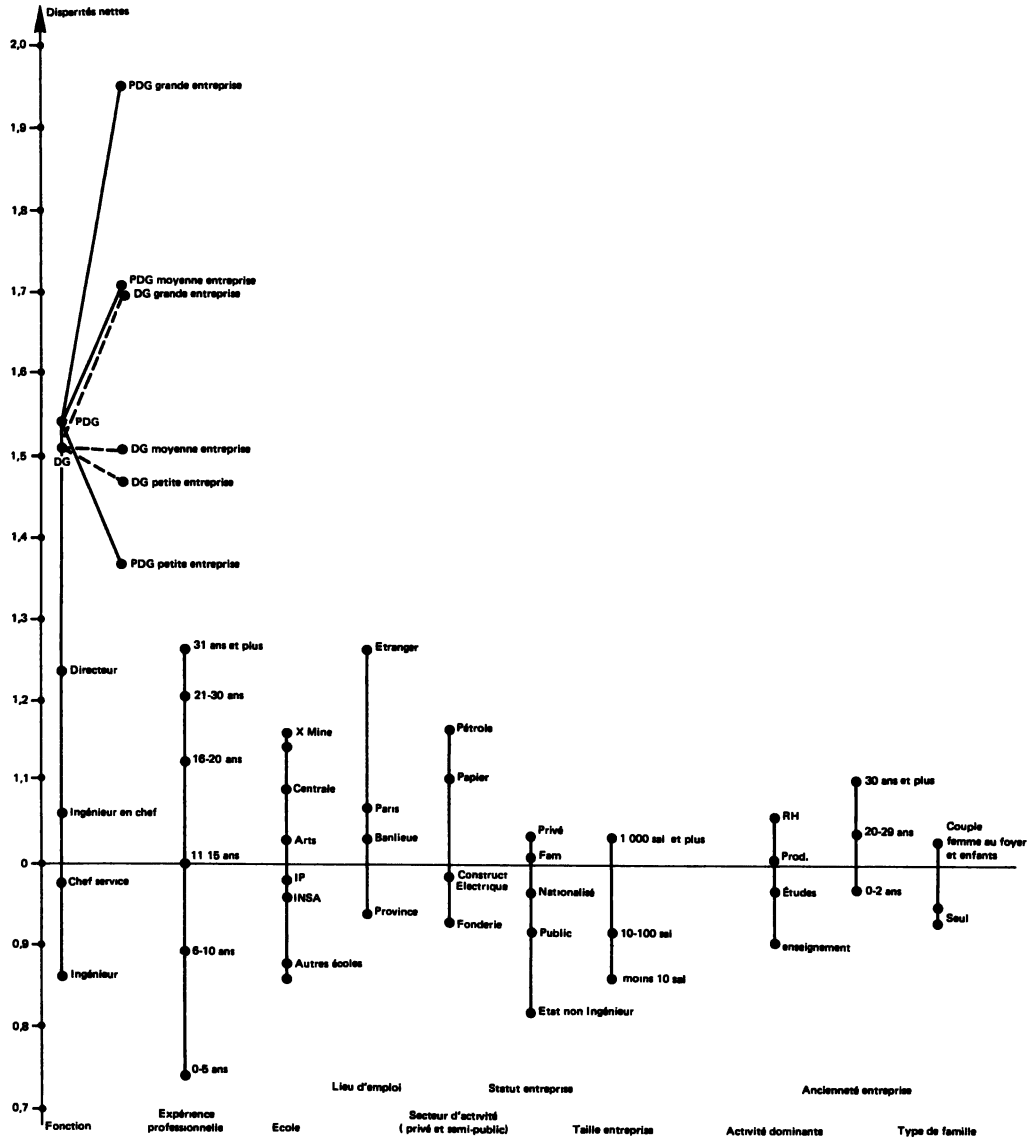
En 1986 la rémunération moyenne brute était de 293 000 F par an, 10 % des ingénieurs percevaient moins de 155 000 F annuels et les 10 % les plus favorisés déclaraient recevoir plus de 470 000 F par an. (La médiane étant à 250 000 F).

C'est en fonction de la position hiérarchique, depuis l'ingénieur de base jusqu'au P-DG salarié, que les écarts de rémunération sont les plus forts, mais aussi que les dispersions s'accroissent (figure 4).

Age ou plutôt expérience professionnelle et école de formation sont aussi des facteurs importants de disparités de rémunération qui mêlent leurs influences aux variables d'entreprise (secteur, statut, taille) et aux autres variables plus caractéristiques du marché du travail (région d'emploi, ancienneté dans l'entreprise).

Au total dix variables ont été retenues dans l'analyse de variance (voir tableau 1), avec une spécification multiplicative (c'est-à-dire additive en logarithme). Près de 70 % de la variance des rémunérations est expliquée par le modèle.

Figure 5. — Disparités nettes de salaire des ingénieurs masculins en 1986*



* Source : Enquête FASFID 1987.

La position hiérarchique et l'expérience professionnelle présentent la plus forte contribution spécifique à l'explication (Rappelons que cette contribution est mesurée par la perte de variance expliquée quand on supprime ce critère de la liste des variables explicatives). Au deuxième niveau, l'école de formation et le lieu de travail (distinguant Paris centre, la région parisienne, la province et l'étranger) présentent un fort pouvoir explicatif. Puis interviennent les variables d'entreprise (secteur, statut, taille). Notons que l'ancienneté est fort peu explicative des disparités de rémunération des ingénieurs diplômés.

Ces résultats se retrouvent à la lecture des coefficients estimés. Ainsi être P.D.G. de grande entreprise (+ 500 sal.) fait presque doubler le niveau de rémunération toutes choses égales par ailleurs... (figure 5).

1. *L'effet qualité ou le rôle de l'école de formation*

Pour illustrer l'importance des effets de structure prenons le cas des disparités de rémunération entre écoles d'ingénieurs.

Si 100 désigne le niveau moyen, les disparités apparentes (peu différentes du rapport des moyennes) sont au niveau 132 pour l'école polytechnique contre 72 pour l'école des techniques industrielles et des Mines d'Alès.

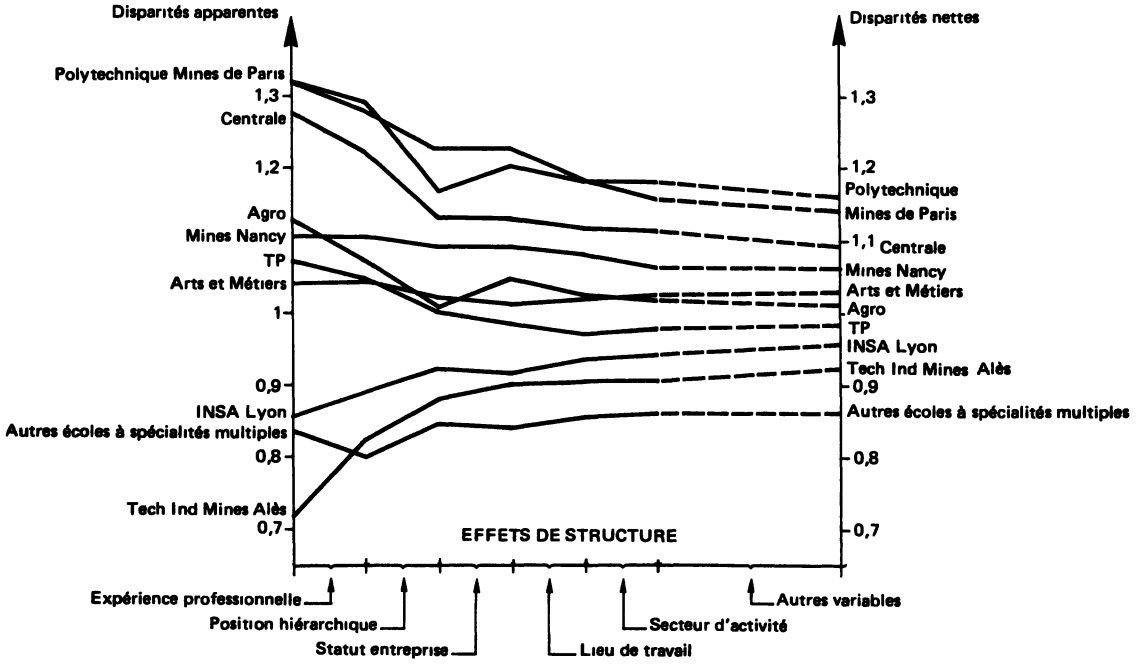
Après correction des effets de structure l'école polytechnique se situe à l'indice 116 et l'école d'Alès à l'indice 92,5 (figure 6). Une grande partie de l'écart est résorbé par les effets de l'expérience professionnelle et de la position hiérarchique. Ainsi 62 % des sortants de l'école d'Alès sont simplement ingénieurs, leur expérience professionnelle moyenne est de 7,7 années contre 17,1 années pour les anciens polytechniciens qui n'occupent des positions d'ingénieur de base que dans 20 % des cas.

En fait, l'effet de structure dû à l'expérience professionnelle exprime les différences de pyramide des âges des anciens élèves par école de formation selon que l'école est ancienne ou plus récente, que ses promotions sont stabilisées ou en forte expansion...

Bien que l'écart des disparités résiduelles entre l'école polytechnique et l'école des Mines de Paris ne soit significative qu'à un seuil dépassant les 10 %, le jeu des effets de structure diffère notablement d'une école à l'autre. Alors que les pyramides d'âge sont comparables, les anciens polytechniciens occupent plus fréquemment que les anciens mineurs des positions hiérarchiques élevées : 16 % d'entre eux sont P.D.G. ou Directeur général, dont 3 % chef d'une entreprise de plus de 500 salariés contre respectivement 11 % et moins de 1 % pour les anciens mineurs. Néanmoins cet avantage qui explique près de 5 points d'écart entre rémunérations est compensé pour les anciens élèves de l'école des mines de Paris

Figure 6

Des disparités apparentes aux disparités nettes selon les écoles d'ingénieurs*



* Source : Enquête FASFID 1987.

par le fait qu'ils n'occupent pratiquement jamais des postes de fonctionnaires non ingénieurs — ces postes relativement moins biens payés (plus de 20 % d'écart avec le secteur privé) sont occupés par 11 % d'anciens polytechniciens.

III. Une approche des profils de carrière

L'étude des carrières réelles de différentes catégories de salariés ne peut se faire que sur des données de Panel [4] ou tout au moins de pseudo-panels (enchaînement de coupes instantanées permettant de repérer sans ambiguïté à chaque date une cohorte d'individus) [3].

L'analyse de variance effectuée sur coupe instantanée permet sous certaines hypothèses d'obtenir des estimations de profils de carrière [1]. Désignons sous ce terme ce qui dans les carrières réelles relève des évolutions liées au vieillissement et aux promotions sans prendre en compte ce qui tient aux évolutions conjoncturelles, habituellement mesuré par les indices de salaire à structure constante. Dans le cas des fonctionnaires les deux dimensions précédentes s'identifient facilement : pour la seconde, à l'augmentation de la valeur du point fonction publique et pour les profils de carrière, à l'évolution des individus dans la grille hiérarchique. Pour les salariés du secteur privé la distinction carrière-conjoncture est moins immédiatement appréhendable mais mérite, nous semble-t-il, toutefois d'être effectuée.

En nous inspirant de la théorie du capital humain pour laquelle le salaire rémunère un stock de capital humain acquis par l'individu avant son entrée sur le marché du travail (formation scolaire) mais qu'il continue d'accroître tout au cours de sa vie active (« Learning by doing »), nous distinguerons deux formes d'accumulation post-scolaire :

- l'une plus générale qui pourra être valorisée sur le marché du travail supposé concurrentiel (le marché externe) et qui sera mesurée par l'expérience professionnelle [2].
- l'autre plus spécifique qui correspond aux investissements réalisés dans l'entreprise et ne pourra être valorisée que sur ce qu'on a appelé le « marché interne ». L'ancienneté dans l'entreprise constituera une mesure de l'intensité de ces investissements spécifiques.

1. *Une rentabilité spécifique du marché interne largement positive*

Estimés sur l'ensemble des salariés masculins à partir de l'enquête sur l'emploi, les avantages associés au marché interne représentent près de la moitié de la rentabilité de l'expérience professionnelle.

Les bénéfices liés à l'expérience professionnelle croissent rapidement en début de vie active au rythme d'environ 1,5 % par an, passent par un maximum vers 25-30 ans de carrière et décroissent par la suite. L'expérience professionnelle est donc un atout en début de carrière et un handicap en fin de carrière (figure 7 A).

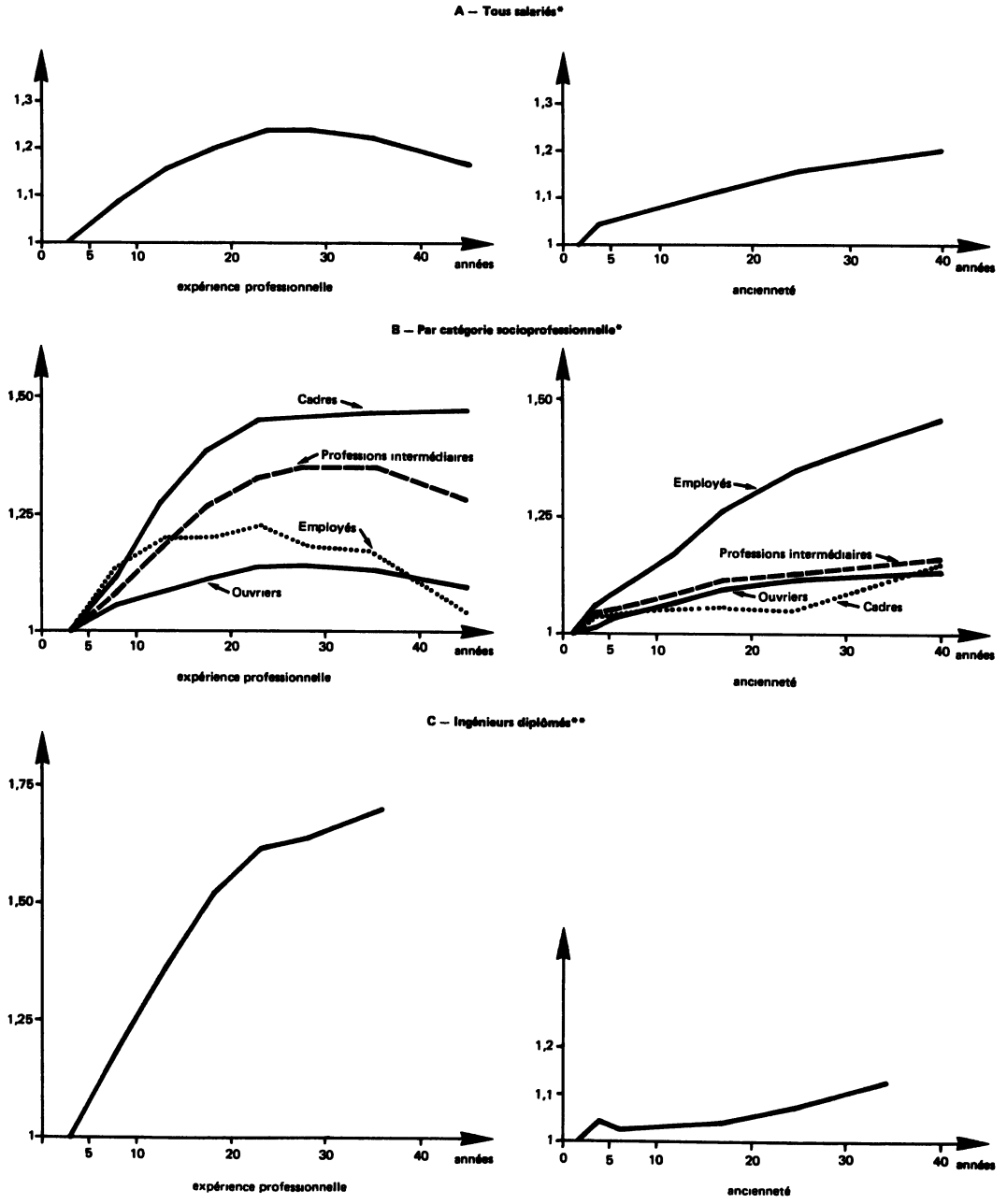
Les avantages spécifiques de l'ancienneté sont aussi plus importants en début de carrière interne, mais restent positifs sur l'ensemble de celle-ci. Évalués à environ 0,8 % par an pour les dix premières années et ils tombent pour les années suivantes à 0,4 % par an.

L'étude des rentabilités comparées des marchés externe et interne peut être approfondie selon les catégories socio-professionnelles en introduisant dans le modèle d'analyse de variance des interactions

1. Ces hypothèses concernent plus particulièrement l'absence de phénomènes de génération et la constance des indices de salaires catégoriels par âge (voir [2] et [3]).

2. Ne disposant pas à l'enquête sur l'emploi et à l'enquête FASFID d'une mesure précise de cette expérience professionnelle totale nous l'estimerons par différence entre l'âge à l'enquête et l'âge à la première embauche, en faisant l'hypothèse que les interruptions d'activité sont négligeables; ce qui est admissible pour les salariés masculins auxquels nous nous restreindrons dorénavant.

Figure 7
Effets spécifiques de l'expérience professionnelle et de l'ancienneté dans l'entreprise sur les salaires masculins



* Source : Enquête emploi 1987.
** Source : Enquête FASFID 1987.

catégorie sociale × ancienneté dans l'entreprise et catégorie sociale × expérience professionnelle. L'explication est globalement améliorée (le pourcentage de variance expliquée passe de 67,7 % à 68,4 %) et les différences entre coefficients sont significatives.

La rentabilité de l'expérience professionnelle apparaît d'autant plus forte que la qualification s'élève. D'ailleurs pour les cadres la rentabilité marginale n'est plus négative après 30 ans (figure 7 B). En revanche la rentabilité du marché interne est minimale pour les cadres et maximale pour les exployés et les catégories intermédiaires.

Ces résultats pourraient s'expliquer par une meilleure transférabilité sur le marché « externe » du travail des investissements post-scolaires accumulés par les cadres au cours de leur carrière professionnelle, comme si « savoir encadrer » pouvait plus facilement s'exporter dans une autre entreprise que posséder une qualification technique très spécialisée et dépendante d'une technologie spécifique.

La mesure de la rentabilité de l'expérience professionnelle et de l'ancienneté chez les anciens élèves des écoles d'ingénieurs (enquête FASFID) confirme les observations précédentes. Alors que les bénéfices attribués à l'expérience professionnelle sont spectaculaires, avec une croissance de plus de 3 % par an en début de carrière (figure 7 C) il faut attendre 25 années de maison pour enregistrer un bonus significatif d'ancienneté. Ainsi, les ingénieurs diplômés présenteraient les plus forts rendements généraux des investissements post-scolaires et les plus faibles rendements spécifiques. Un haut niveau de formation initiale semble être le gage d'une meilleure rentabilité des apprentissages sur le tas.

La validité des résultats précédents reste néanmoins conditionnée par la nature du modèle utilisé. Premièrement, les analyses sur coupe instantanée peuvent être biaisées par des effets de génération plus ou moins bien contrôlés. Deuxièmement, la mesure des effets de carrière tant sur le marché externe que sur le marché interne est effectuée à qualification constante (en termes de catégorie socioprofessionnelle à deux chiffres). Ainsi, toute promotion salariale accompagnée d'un changement de catégorie socioprofessionnelle n'est pas comptée comme effet de carrière. Enfin, le fait que la fonction exercée par les cadres figure explicitement comme variable d'analyse réduit d'autant le pouvoir explicatif des variables de carrière et tout particulièrement de l'ancienneté dans l'entreprise. Là encore, notre mesure des rentabilités de l'expérience professionnelle et de l'ancienneté est faite à fonction ou position hiérarchique constante, et ne correspond pas forcément à l'idée commune de carrière, pour laquelle tous ces facteurs se mêlent et cumulent leurs effets.

Plus généralement, les modèles multivariés fournissent des résultats d'ordre analytique relevant d'une certaine abstraction. Et le fameux « toutes choses égales par ailleurs » qui revient comme un leitmotiv dans les commentaires n'est pas toujours très facile à bien interpréter, même pour les spécialistes, et ceci pour deux raisons principales.

Tout d'abord, la réalité sociologique démontre assez que les états du système ne sont jamais équidistribués et que bien souvent les « effets de structure » tendent à cumuler leurs avantages. Ainsi, le lecteur moyen aura des difficultés à comprendre que l'on veuille séparer les effets grande entreprise, région parisienne, proportion de cadres... sur les salaires, alors que ces facteurs vont de pair dans la réalité et forment plutôt un pôle de « bons salaires ». D'où en particulier l'intérêt et le succès des méthodes d'analyse factorielle (analyse des correspondances et en composantes principales), beaucoup plus synthétiques et qui conduisent à des typologies plus facilement accessibles au lecteur.

La seconde limite à la bonne interprétation des résultats est plus fondamentale et tient à l'enregistrement des données. Plusieurs exemples nous ont montré à quel point dire que tel critère est plus explicatif que tel autre dépend de la finesse de la nomenclature retenue et de la présence ou de l'absence d'autres variables explicatives. Quelle que soit la spécification choisie, le modèle n'explicitera jamais que ce que le statisticien ou ses prédécesseurs auront identifié et séparé dans des codes et nomenclatures. La mesure statistique est donc largement déterminée par les catégories en usage dans la représentation du monde social. Gageons qu'en retour elle participe aussi à son élaboration.

TABLEAU 2
LES DOUZE FACTEURS QUI DÉTERMINENT VOTRE SALAIRE EN 1987

Source : Enquête sur l'emploi de mars 1987

CRITÈRES	Hommes	Femmes
1. Votre fonction		
Cadres fonction publique	1,454	1,560
Professeurs	1,459	1,594
Professions information, arts, spectacles	1,366	1,363
Cadres des entreprises Direction générale	2,027	—
Financière, commerciale	1,671	1,765
Autre	1,572	1,600
Ingénieurs et cadres techniques .		
Direction générale, financière	1,880	—
Commerciale, production, entretien	1,663	—
Études, informatique	1,622	1,647
Autre	1,533	1,585
Instituteurs	1,095	1,247
Professions intermédiaires santé	1,060	1,116
Professions intermédiaires fonction publique	1,165	1,167
Direction générale, financière	1,248	1,279
Commerciale	1,276	1,358
Autre	1,124	1,192
Techniciens études, informatique	1,168	1,221
Autre	1,118	1,175
Contremaîtres production, entretien	1,201	1,125
Autre	1,105	1,195
Employés fonction publique	0,862	0,946
Policiers militaires	0,997	1,084
Employés administratifs	0,961	1,001
Employés de commerce	0,885	0,903
Personnel de service	0,799	0,708
OQ industriel	0,912	0,893
OQ artisanal	0,890	0,894
Chauffeurs	0,868	0,910
OQ manutention transport	0,863	0,902
ONQ industriel	0,822	0,836
ONQ artisanal	0,812	0,791
Salarié agricole	0,767	0,873
2. Vos diplômes		
Pas d'études	0,874	0,909
Pas de diplôme	0,921	0,897
CEP	0,948	0,918
BEPC seul	1,016	1,000
CAP	0,991	0,982
BEPC + CAP ou FPA	1,022	1,011
BP, BEI, BAC technique	1,054	1,062
BAC 2e partie	1,069	1,069
DUEL, DUT, santé..	1,112	1,134
Supérieur (nc Grandes écoles)	1,260	1,188
Grandes écoles	1,458	1,262
3. Votre ancienneté dans l'entreprise		
0 2 ans	0,938	0,894
3-4 ans	0,961	0,963
5 6 ans	0,975	0,972
7 9 ans	0,989	0,985
10-14 ans	1,008	1,024
15 19 ans	1,042	1,059
20 29 ans	1,065	1,137
30 et plus	1,121	1,197
4. Votre durée hebdomadaire de travail		
30-34 heures	0,943	0,864
35 37 heures	0,973	0,972
38 heures	0,978	1,020
39 heures	0,967	1,000
40 heures	0,998	1,017
41 43 heures	1,036	1,044
44-47 heures	1,070	1,044
48 52 heures	1,128	0,993
53 heures et plus	1,174	0,999
5. L'activité de votre établissement		
Agriculture	0,955	0,980
Viande, lait	0,965	0,998
Autres IAA	1,040	1,028
Combustibles solides	1,006	1,153
Pétrole, gaz naturel	1,191	1,128
Électricité, gaz et eau	1,040	1,035
Métaux ferreux, acier	0,990	1,017
Métaux non ferreux	1,075	1,103
Matériaux de construction	0,982	1,012
Verre	1,071	1,039
Chimie	1,101	1,130
Parachimie, pharmacie	1,108	1,125
Fonderie	1,014	1,009
Construction mécanique	1,018	1,064
Construction électrique et électronique	1,030	1,071
Matériel de transport terrestre	0,985	1,087
Construction navale et aéronautique	1,066	1,084
Textile, habillement	0,962	0,956
Cuir, chaussure	0,944	1,004
Bois, meubles	0,971	1,021
Papier, carton	1,100	1,041
Imprimerie, presse, édition	1,147	1,109
Caoutchouc, plastiques	1,025	1,049
Bâtiment, génie civil et agricole	0,989	1,000
Commerce gros alimentaire	1,004	1,009
Commerce gros non alimentaire	1,016	1,040
Commerce détail alimentaire	0,921	0,927
Commerce détail non alimentaire	0,967	0,982
Réparation automobile	0,984	0,966
Hôtels, cafés, restaurants	0,962	1,132
Transports	1,033	1,040
Télécommunications, postes	0,937	0,948
Services marchands aux entreprises	1,018	1,013
Services marchands aux particuliers	0,991	1,028
Location, crédit bail	0,936	0,953
Assurances	1,019	1,007
Organismes financiers	1,054	1,113
Services non marchands	0,965	0,933
6. Le statut juridique de votre entreprise		
Secteur privé, autre salarié	0,997	0,987
Secteur privé, salarié intérimaire	0,969	1,049
Secteur privé, salarié saisonnier	0,979	0,987
Secteur privé, contrat à durée déterminée	0,951	0,913
Secteur privé, salarié d'un parent	1,011	0,950
Secteur privé, salarié à l'essai	0,905	1,059
État	1,019	1,030
Collectivités locales, hôpitaux, HLM	0,990	1,016
Sécurité sociale	1,033	1,058
Entreprise publique	1,027	1,013
7. La taille de votre établissement		
1 2 salariés	0,935	1,008
3-4 salariés	0,964	1,018
5 9 salariés	0,974	1,001
10 19 salariés	0,976	1,008
20 49 salariés	0,994	1,001
50 99 salariés	0,997	0,985
100-499 salariés	1,010	1,014
500 999 salariés	1,017	1,026
1 000 salariés et plus	1,033	1,012
8. Votre expérience professionnelle totale		
0 5 ans	0,869	0,908
6 10 ans	0,938	0,969
11 15 ans	0,990	1,009
16 20 ans	1,027	1,035
21 25 ans	1,051	1,031
26 30 ans	1,054	1,050
31 40 ans	1,043	1,023
41 ans et plus	1,007	1,002
9. Votre position dans le ménage		
Personne vivant seule	0,968	1,022
Chef ou conjoint d'un couple :		
sans enfant, femme active	0,991	1,010
avec enfant (1), femme active	1,012	0,995
sans enfant, femme au foyer	1,036	—
avec enfant (1), femme au foyer	1,043	—
chef d'un ménage sans couple	0,975	1,015
Autres positions (enfant .)	0,921	0,958
10. Votre nationalité		
France	1,002	1,001
Afrique	0,941	0,980
Italie, CEE	1,067	0,984
Espagne, Portugal	1,015	0,955
Pologne, Turquie, Yougoslavie	0,943	1,011
Autres	0,936	1,084
11. Votre région de résidence		
Ile de France	1,084	1,073
Champagne Ardennes	0,965	0,992
Picardie	0,991	0,998
Haute Normandie	0,932	0,970
Centre	0,982	0,970
Basse Normandie	0,965	0,952
Bourgogne	0,961	0,971
Nord Pas-de Calais	0,936	0,920
Lorraine	0,977	0,978
Alsace	1,046	1,001
Franche-Comté	0,965	0,970
Pays de Loire	0,953	0,964
Bretagne	0,969	0,956
Poitou Charentes	0,949	0,943
Aquitaine	0,970	0,957
Midi Pyrénées	0,952	0,951
Limousin	0,943	0,941
Rhône Alpes	1,011	1,010
Auvergne	0,932	0,972
Languedoc Roussillon	0,959	0,950
Provence-Côte d'Azur-Corse	1,016	1,002
12. Votre commune de résidence		
Commune rurale d'un canton rural	0,985	0,986
Commune rurale d'un canton partiellement urbain	1,000	0,991
Unité urbaine de moins de 20 000 habitants	0,985	0,976
Unité urbaine de 20 000 à moins de 100 000 habitants	0,985	0,994
Unité urbaine de 100 000 habitants et plus	1,005	1,005
Agglomération parisienne hors Paris	1,018	1,013
Paris ville	1,046	1,044

Les données figurant dans ce tableau sont les disparités ou effets spécifiques correspondant aux différentes variables explicatives des salaires en 1987. Par exemple, les hommes cadres de la fonction publique gagnent, toutes classes égales par ailleurs, 1,454 fois le salaire moyen masculin en 1987.

ANNEXE

L'ANALYSE DE LA VARIANCE

Soit le modèle suivant où i désigne l'individu, W_i son salaire mensuel X_j, Y_k, Z_l... les caractéristiques sociodémographiques.

$$Log\, W_i = cte + \sum_j a_j X_{ij} + \sum_k b_k Y_{ik} + \sum_l c_l Z_{il} + \ldots + u_i$$

$$\text{avec } u_i \approx N(0,s^2)$$

Si par exemple X_j désigne la catégorie socioprofessionnelle, j aura 23 valeurs et la aj seront les estimateurs des disparités nettes ou effets spécifiques associés à la variable catégorie socioprofessionnelle. Pour une liste détaillée des modalités de chaque variable de l'analyse on se reportera au tableau 2. Les valeurs des paramètres de chaque effet ont été calculées de façon à ce que la constante corresponde à la moyenne de la variable expliquée soit $\overline{\text{Log } W_i}$ ($\neq \log\overline{W_i}$) et à ce que l'effet moyen de chaque variable soit nul.

Disparité apparente

Les disparités de salaire entre catégories sont habituellement appréhendées par le rapport des moyennes. Parfois le rapport des médianes est utilisé pour éviter que des valeurs extrêmes ne perturbent le calcul de la moyenne. Ici nous avons retenu l'exponentielle de l'écart entre la moyenne des logarithmes de salaires.

Notons que l'indicateur retenu se rapproche du rapport des moyennes si les deux catégories ont la même dispersion. On appellera disparité apparente pour une sous-population la valeur de l'indicateur de disparité précédemment défini lorsque la deuxième sous-population correspond à l'ensemble de la population. La disparité apparente est donc proche du rapport de la moyenne de la sous-population à la moyenne générale.

Des disparités apparentes aux disparités nettes

En sommant les valeurs de l'ajustement précédent pour les n_j salariés correspondant à la modalité j du critère X on obtient la relation suivante :

$$exp\left(\frac{1}{n_j}\sum_{i \in (X_j = 1)} Log\, W_i - \overline{Log\, W_i}\right) = exp\, a_j \cdot exp\left(\sum_k \frac{n_{jk}}{n_j} b_k\right) \cdot exp\left(\sum_l \frac{n_{jl}}{n_j} c_l\right).$$

Le terme de gauche est la disparité apparente associée à la modalité j du critère X, exp a_j est la disparité nette ou effet spécifique et on passe de la disparité nette à la disparité apparente par produit d'un certain nombre d'effets de structure dus aux autres variables de l'analyse.

Quand le critère étudié, comme l'âge par exemple, n'appartient pas aux variables de l'analyse, le terme correspondant à la disparité nette (exp aj) est remplacé par la moyenne des résidus pour les salariés appartenant à la modalité m du critère étudié. Ceci permet d'étudier la capacité du modèle

à rendre compte des disparités de salaire pour un critère ne figurant pas parmi les variables explicatives, en passant cette fois-ci des disparités apparentes aux résidus non expliqués par le modèle.

La bonne interprétation des disparités nettes

L'interprétation des disparités régionales ne suscite guère de controverse car si chaque région comporte une structure particulière de qualifications, d'emplois, de secteurs d'établissements... les spécificités régionales ne sont pas telles que tous les cadres habitent en région parisienne ou que toute l'automobile soit concentrée en Franche-Comté. Il en est de même pour l'estimation d'un effet nationalité ou taille d'entreprise. L'interprétation de l'estimation des disparités nettes dues à la qualification et au diplôme pose des questions plus délicates.

Cela ne revient-il pas à estimer une différence de salaire entre cadres et ouvriers à structure de diplômes constante comme si on pouvait trouver des ouvriers et des cadres à chaque niveau de l'éventail des diplômes? Remarquons tout d'abord que si tous les ouvriers avaient le CAP et tous les cadres étaient diplômés de grande école, on ne pourrait estimer séparément un effet diplôme et un effet qualification, les deux variables étant parfaitement corrélées. Ce n'est que parce qu'il existe une distribution non diagonale de population entre qualification et diplôme que le modèle peut fonctionner, en réalité parce que si les ouvriers diplômés sont rares, les cadres autodidactes sont nombreux (40 % des cadres administratifs et commerciaux des entreprises n'ont pas le baccalauréat!).

Le chameau au pôle Nord et le renne au Sahara

Mais que signifient alors les résultats? Détaillons notre petit exemple précédent avec deux qualifications et deux diplômes et supposons que tous les ouvriers ont le CAP et que les cadres se répartissent entre titulaires du CAP et diplômés de grande école. Dans ce cas l'effet diplôme sera estimé par l'écart de salaire entre cadre diplômé et cadre non diplômé et l'effet qualification correspondra à l'écart entre le salaire moyen ouvrier et le salaire moyen des cadres non diplômés. Cet exemple prouve que le commentaire des résultats d'une telle analyse doit, en cas de liaison forte entre deux critères, être replacé dans le cadre de cette liaison, les disparités nettes devenant ici des effets « très spécifiques ».

Remarquons que si quelques OS diplômés de grande école (rennes au Sahara ou chameaux au pôle Nord selon l'expression de SIMIAND) s'étaient glissés dans l'échantillon cela n'aurait eu qu'une influence minime sur les estimations, même si la qualité de l'ajustement pour ces individus était particulièrement mauvaise (mais l'ambition de ces modèles n'est pas de bien représenter les cas trop particuliers).

La qualité du modèle utilisé

Toutes les estimations calculées par le modèle reposent sur l'hypothèse que l'on a bien retenu l'ensemble des variables explicatives pertinentes sans en omettre. Ou plutôt que les variables omises sont indépendantes des variables incluses, ce qui n'est pas toujours le cas. Ainsi pour reprendre l'exemple de la liaison formation-qualification, la suppression de l'une de ces variables dans l'estimation provoque une modification importante des valeurs des estimateurs de la variable restante. A cette dernière s'ajoute en quelque sorte l'influence de la variable omise par l'intermédiaire des corrélations entre variables. En supprimant la formation, l'écart entre OS et cadre s'accroîtra, la modalité cadre incorporant alors en quelque sorte l'effet de qualification et de diplôme (car les cadres sont nombreux à être diplômés).

Dans cette étude, le pouvoir explicatif spécifique de chaque critère est mesuré par la perte en explication totale (pourcentage de variance expliquée) occasionnée par la suppression du critère dans

la liste des variables explicatives. Comme on l'a vu, les autres critères les plus corrélés à la variable supprimée peuvent compenser partiellement cette absence. Le pouvoir explicatif spécifique sera ce qui ne pourra pas être totalement compensé.

L'hypothèse d'additivité des effets de chaque variable sur le logarithme du salaire doit également être appréciée pour assurer la pertinence du modèle utilisé. Nous le ferons en examinant l'importance et la régularité des écarts résiduels entre valeur vraie et valeur estimée pour des sous-populations obtenues par croisement des critères deux à deux.

BIBLIOGRAPHIE

[1] M. GLAUDE : « La structure des salaires en 1985 ». Données Sociales 1987.

[2] M. GLAUDE : « Ancienneté, expérience et théorie dualiste du marché du travail : une étude sur données individuelles ». Économie appliquée n° 4, 1986.

[3] Ch. BAUDELOT et M. GLAUDE : « Étude de la relation Salaire formation. Un essai d'analyse longitudinale à travers trois enquêtes transversales ». Communication présentée au VIIIe journées d'économie sociale. Dijon 1988. A paraître dans Économie et Statistique.

[4] S. LOLLIVIER et J.F. PAYEN : Les carrières salariales des hommes entre 1967 et 1982. A paraître dans Données Sociales 1990.