

**UNIVERSITE DE PARIS X-NANTERRE
U.F.R. DE SCIENCES ECONOMIQUES**

DOCTORAT

**ANALYSE SPATIO-TEMPORELLE PAR LA MODELISATION
LOG-LINEAIRE :**

Une approche structurelle de l'offre de l'emploi salarié en France.

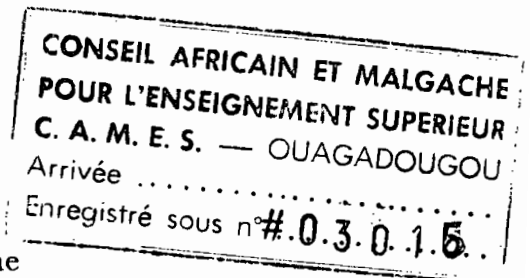
Présentée et Soutenue Publiquement

par

Georges KOBOU

Directeur de recherche
Professeur à l'Université de Paris X-Nanterre

O. FAVEREAU,



#03016

JURY :

G. ABRAHAM-FROIS, Professeur à l'Université Paris X, Président

P. COHENDET, Professeur à l'Université Louis-Pasteur Strasbourg,
Rapporteur

O. FAVEREAU, Professeur à l'Université Paris X, Rapporteur

J.A ZIGHERA, Professeur à l'Université Paris X

L'Université de Paris X-Nanterre n'entend donner aucune approbation ou improbation aux opinions émises dans les mémoires : ces opinions doivent être considérées comme propres à leur auteur.

AVANT-PROPOS

Je tiens à remercier :

- M. Le Professeur J.A. ZIGHERA et M. Le Professeur O. FAVEREAU pour leur grande disponibilité, ainsi que leurs critiques, conseils et suggestions, lesquels auront contribué, à l'orientation donnée et à l'amélioration décisive de cette recherche. Qu'ils trouvent ici l'expression de ma très profonde reconnaissance ;
- M. Le Professeur G. ABRAHAM-FROIS et M. Le Professeur P. COHENDET, de l'honneur qu'ils me font d'être membres de jury ;
- M. MBIADJEU TCHAMBA et Mme Maffo NGONGANG, pour leurs soutiens moral et matériel constants ;
- Mme A. NGONGANG, Mlle A. KAMTA, M. NJEMEN et M. NGOMOU qui ont significativement contribué à l'élaboration de ce travail ;
- Mes amis, particulièrement toute l'aimable et sympathique équipe du Centre d'Analyses Statistiques des Structures et des Flux (CASSF) de l'Université Paris X-Nanterre, dont les remarques et le soutien logistique m'ont permis de mener ce travail à son terme ;
- Tous ceux qui, de près ou de loin, se sont impliqués dans la réalisation de cette thèse.

A MA MERE

«Toutes les connaissances que nous possédons dépendent des méthodes par lesquelles nous distinguons le semblable du dissemblable. Plus le nombre de distinctions que permettent ces méthodes est grand, plus notre connaissance des choses s'accroît. Plus les objets qui nous intéressent sont nombreux, plus il est difficile de créer une telle méthode, mais plus cela est nécessaire».

(LINNAEUS, *Génera Plantarum*, 1737)

SOMMAIRE

INTRODUCTION GENERALE

Chapitre I :

RAPPELS THEORIQUES ET EMPIRIQUES DE L'EVOLUTION GLOBALE DU
MARCHE DU TRAVAIL EN FRANCE DE L'APRES-GUERRE JUSQU'A LA CRISE

Chapitre II :

APPROCHE SPATIALISEE DU MARCHE DU TRAVAIL

Chapitre III :

CONCEPTS DE BASE NECESSAIRES A L'ANALYSE
DU MODELE LOG-LINEAIRE

Chapitre IV :

MODELISATION LOG-LINEAIRE : FORMULATION DE BASE ET SON EXTENSION
PAR LE BIAIS DU CENTRAGE DES PARAMETRES.

Chapitre V :

MODELISATION LOG-LINEAIRE DES TABLEAUX DE CONTINGENCE
AYANT DES MODALITES ORDINALES

Chapitre VI :

PRESENTATION DES DONNEES ET ANALYSE DES
RESULTATS STATISTIQUES

Chapitre VII :

ANALYSE STRUCTURELLE ET DYNAMIQUE DE L'EMPLOI
SALARIE ENTRE 1976 ET 1981

CONCLUSION GENERALE.

INTRODUCTION GENERALE

La crise amorcée dès 1974 a profondément et durablement affecté l'activité économique qui avait connu une croissance remarquable durant les années de l'après-guerre. La baisse de l'activité a provoqué de nombreux déséquilibres dont l'ampleur est manifeste sur le marché du travail. En effet, on y observe en permanence des tensions induites par la divergence entre les tendances de la population active (elles sont amplifiées par l'arrivée de nombreuses générations nées de l'après-guerre, les comportements nouveaux notamment le développement rapide et irréversible de l'activité féminine etc.), et celles du rythme de création des emplois. Au total, l'apport moyen en main-d'œuvre est d'environ 250 000 par an sur le marché du travail, soit presque 1,7 million de personnes sur la période 1975-1982 [1]¹.

0.1 La conception traditionnelle du marché du travail à l'épreuve des faits.

Le comportement atonique de l'emploi n'a donc pas permis de préserver une situation d'équilibre sur le marché du travail où le phénomène majeur est la montée régulière et non contenue du chômage : le nombre de personnes au chômage a plus que doublé entre 1975 et 1982, nonobstant toute la panoplie de mesures mises en place par les pouvoirs publics pour infléchir sa tendance et le juguler. Le caractère inopérant de ces mesures a conduit les économistes du travail à revoir les théories dont elles s'inspirent. Ces théories sont centrées sur les conceptions traditionnelles du système économique en général, ou du système de l'emploi en particulier, et ne parviennent plus à rendre fidèlement compte des mécanismes qui les sous-tendent. Ainsi, les déséquilibres de l'emploi semblent insensibles aux "remèdes macro-économiques" habituels, qui ont pourtant prouvé leur efficacité durant les trente années de l'après-guerre. L'échec des politiques d'inspiration keynésienne mises en œuvre en France sur les périodes récentes (1975-1976 et 1981-1982) est à cet égard symptomatique. Le chômage n'est plus seulement le reflet des déséquilibres globaux ; ses aspects sectoriels, structurels, voire régionaux peuvent même être dominants.

L'analyse du dysfonctionnement du marché du travail ne constitue pas le point central des investigations que nous entreprendrons dans la présente thèse. Aussi, est-il nécessaire de souligner que les faits que nous venons de rappeler rapidement ne servent qu'à montrer comment

¹ Les numéros en [] renvoient aux notes en fin de chaque chapitre.

l'évolution du système économique, auquel le marché du travail participe, produit la mutation de celui-ci. Il n'est donc pas surprenant que les outils traditionnels d'analyse des mécanismes du fonctionnement du marché du travail, élaborés dans un autre contexte précis, ne soient plus aptes à identifier certains phénomènes nouveaux qui y apparaissent, et qui entrent dans le cadre de ce que J. VINCENS a appelé «les nouveaux aspects du problème de l'emploi»[2]. J.P FITOUSSI mentionnera d'ailleurs que «certains aspects du problème de l'emploi sont nouveaux en raison des caractéristiques récentes spécifiques au marché du travail»[3]. Certes, ce problème a été constamment abordé dans de nombreuses monographies y relatives au cours les deux dernières décennies, mais la question de fond qui reste posée est celle de savoir, si les instruments d'analyse proposés permettent de mieux le cerner afin d'enrichir les réflexions théoriques.

Les difficultés rencontrées par les approches macro-économiques pour traduire les phénomènes nouveaux du marché du travail ont conduit à la remise en cause du modèle traditionnel. De nombreuses études empiriques, suscitées assez souvent par les nécessités de l'action politique, avaient par ailleurs déjà mis en lumière les premières critiques de ce modèle, qui a longtemps préconisé l'existence d'un marché unique, présenté comme le lieu d'une confrontation entre une offre et une demande de travail indépendantes, qui aboutit à la détermination d'un taux de salaire et un niveau d'emploi d'équilibre. Et à la suite des travaux de P. DOERINGER et M. PIORE, le caractère fragmenté du marché du travail a été mis en évidence, de même que l'hétérogénéité des emplois et les actifs susceptibles de les occuper.

On se situe alors dans la perspective des théories de la segmentation, qui établissent les phénomènes de diversité des emplois suivant les caractéristiques des actifs (sexe, âge, qualification, métier, etc.) et suivant celles des entreprises (activité principale, politique de gestion du personnel, localisation géographique, taille, etc.). Le succès que connaît ce courant, est lié à la quête quasi-générale ces dernières années, d'une nouvelle problématique en vue d'une reconstitution théorique susceptible de se substituer au modèle traditionnel, qui s'adapte de plus en plus difficilement à la réalité.

Notre contribution dans la présente recherche n'a pas la prétention de proposer un cadre théorique élaboré pour l'étude des mécanismes du marché, compte tenu du foisonnement des

facteurs qui y interviennent. L'imbrication et les interrelations entre les facteurs rendent par la même occasion toute étude de son fonctionnement fort complexe. Aussi, nous limiterons-nous uniquement à l'étude du comportement de l'offre d'emploi, qui constitue à n'en pas douter, une des composantes clés du marché du travail, du fait de son rôle fondamental dans la régulation de ce dernier.

Ce rôle est à rapprocher du paradigme selon lequel l'explication des mouvements d'emploi sur ce marché serait à rechercher du côté des entreprises, les comportements des individus ne pouvant être que de réaction ou d'adaptation (cela n'exclut pas qu'il y ait une certaine autonomie des adaptations). Dans le même sens, STOECKEL a souligné le rôle prépondérant des entreprises dans le processus d'allocation de la main-d'œuvre, car dit-il, «l'entreprise est une catégorie significative de la répartition des postes dans la mesure où son niveau technologique et son organisation interne régissent les emplois et donc l'accès aux emplois»[4]. Cette primauté de l'entreprise dans le processus d'allocation du travail est également partagée par P. DOERINGER et M. PIORE, qui estiment que «le marché du travail est structuré par des unités de production et dépend en particulier de l'état de la technologie»[5]. Ainsi, l'étude de l'offre d'emploi est essentielle, puisqu'elle traduit le rôle actif des entreprises dans l'expression de la demande du travail : les théories de la segmentation tentent de donner la traduction de ce constat dans le fonctionnement du marché du travail.

Il nous appartiendra de cerner les limites des théories, surtout du point de vue des outils d'analyse. Dans cette perspective, afin de concilier la théorie et les faits, on pourra voir comment l'utilisation des outils nouveaux pour l'étude des comportements de l'offre d'emploi, permet de dégager quelques principes d'analyse des mécanismes du fonctionnement du marché du travail.

Les recherches actuelles mettent l'accent sur la dimension qualitative de l'emploi. Au-delà, les différenciations observées sur le marché du travail étant considérées comme produites par le système économique lui-même – c'est du moins ce que présupposent les théories de segmentation – et non sur les diversités individuelles, il serait nécessaire de mettre en lumière les "effets de structure" du marché, ce dans la mesure où l'évolution du système économique impose de nouvelles structures. Celles-ci «se collent, se juxtaposent, se mêlent, détruisent des tissus

sociologiques qui ont des histoires et qui réagissent de diverses façons ». C'est ainsi par exemple que la création d'unités productives assez comparables sur deux marchés différents, par les anciennes structures qui les ont caractérisées, peut ne pas avoir les mêmes effets sur le niveau de l'emploi.

0.2 Nécessité d'une approche multidimensionnelle.

Etant donné que l'emploi recouvre un faisceau de relations complexes, où se mêlent et s'imbriquent plusieurs facteurs, nous essaierons de déceler ces structures par son approche multidimensionnelle, qui permet de mettre en relief l'interaction entre les variables : « l'emploi n'est pas seulement la résultante des facteurs économiques et démographiques, il dépend également des comportements sociologiques et même psychologiques qui viennent perturber les schémas de régulation traditionnels autrefois opératoires et implique une approche multidimensionnelle, cherchant à mettre en évidence toutes les interrelations du système de l'emploi »[6]. On voit dès ce niveau apparaître la nécessité d'une analyse multidimensionnelle

BLACKITT notait que «les analyses multidimensionnelles ne donnent pas une réponse à un problème donné, mais fournissent d'une réalité trop complexe pour être directement accessible à notre entendement des images simplifiées, mais aussi peu déformées que possible, permettant chacune de saisir des traits essentiels de la structure générale et de la reconstituer »[7]. Ces analyses constituent des microscopes par lesquels on peut décrypter les structures latentes des tableaux à grands corpus de données, portant sur les variables du marché du travail.

Potentiellement, les données relatives à l'emploi et issues des enquêtes, se présentent donc sous forme des tableaux multidimensionnels, où figure à titre d'exemple le nombre de salariés d'une catégorie d'âge, travaillant dans un secteur d'activité donné et appartenant à une certaine catégorie socioprofessionnelle, à l'intérieur d'un espace géographique bien déterminé. Mais pour des raisons de place et de commodité de lecture, ces données ont été rarement accessibles sous cette forme et on se ramène très généralement à des tableaux à deux dimensions, correspondant à des sommations partielles, et par conséquent à une partie seulement de l'information

potentiellement disponible. Il s'agit alors d'une approche limitée, ne permettant pas d'affiner au mieux les caractéristiques des données associées aux variables du marché du travail.

L'approche multidimensionnelle traduit donc le recours aux analyses déglobalisées du marché du travail, car la plupart des données utilisées sont agrégées et contiennent des grandeurs hétérogènes. Compte tenu des observations préalablement mentionnées, l'étude de l'offre d'emploi sera abordée ici à travers quatre variables dont le croisement conduira aux analyses structurelles, grâce aux techniques instrumentales s'y prêtant fort bien. Il s'agit des variables que sont l'âge, le sexe, la profession, la région. Cette dernière méritera une attention particulière, puisqu'elle nous permet de repérer les modalités d'une intégration de l'espace dans le marché du travail.

0.3 Rapport entre l'espace et le marché du travail.

En effet, pour les responsables politiques, le premier niveau auquel se fait sentir toute difficulté en matière d'emploi est le niveau régional ou local. C'est à ce niveau que se situe la perception du marché du travail par les agents. La définition d'objectifs en matière de créations d'emplois doit donc intégrer la dimension régionale. D'ailleurs, les années de crise, notamment celle de l'emploi, se caractérisent par la multiplicité des initiatives prises au niveau local. Ces quelques aspects mettent en lumière la reconnaissance implicite de l'incapacité du modèle global à expliquer les dysfonctionnements régionaux, voire locaux. C'est afin de parer à une telle insuffisance qu'apparaît l'une des raisons de prendre en compte l'espace dans le marché du travail.

Cependant, la prise en compte de l'espace dans le marché du travail ne doit pas seulement être perçue comme la nécessité de choisir un cadre réduit pour l'analyse des mécanismes du marché en vue de la reconstitution d'un modèle théorique général – la problématique des fondements micro-économiques de la macro-économie est ici sous-jacente – mais également comme celle de situer chaque agent dans un environnement concret qui va peser sur ses décisions et moduler son comportement : le salarié ressent les contraintes du système économique à travers les structures avec lesquelles il entre en relation. Il a de ce point de vue une connaissance

spatialement limitée. D'autre part, l'entreprise est en relation avec un nombre réduit d'agents et l'intensité des relations est liée à l'aire géographique où s'étend son influence. Dès lors, l'intégration de l'espace dans le marché du travail ne permettrait pas d'isoler l'agent économique des relations dans lesquelles il s'insère. Nous touchons ici l'un des points importants des mécanismes du marché, où la dimension spatiale est gommée. Quant à la deuxième hypothèse elle traduit l'hétérogénéité des relations, dont on pourrait faire le parallélisme avec l'ancrage spatial du travail, qui joue le rôle prépondérant de facteur structurant l'espace. Celui-ci est alors hétérogène du point de vue du facteur travail. Mais étant donné que l'évolution du système économique produit la mutation du marché du travail, lequel a un rapport certain à l'espace, il serait indispensable d'analyser cette hétérogénéité à travers le temps, afin de mieux appréhender les évolutions structurelles. Dans une optique plus globale, N. DUBRULLE met l'accent sur cette approche dynamique dans la mesure où précise-t-elle «la théorie économique ne peut rendre compte de la réalité dans toute sa complexité et les données macro-économiques [...] nous questionnent néanmoins sur l'adéquation au réel et la permanence des mécanismes formalisés par rapport à l'hétérogénéité spatiale et temporelle du comportement des agents [...]. C'est pourquoi l'analyse des paramètres économiques et sociaux à un niveau fin [...] reste indispensable pour une analyse dynamique en termes complexes » [8].

Notre travail se situe également dans une perspective dynamique, l'exploration des structures de l'emploi s'effectuant dans un intervalle de temps donné. Bien entendu, le décryptement de ces structures se fera par le biais d'un outil statistique *sui generis*, qui de surcroît s'accommode par parfaitement aux analyses descriptives.

0.4 L'intérêt d'une analyse descriptive.

La crise du système économique en général ou celle de l'emploi en particulier, ne se traduit pas seulement par une variation des agrégats, mais aussi par une variation de leur composition. Il s'agit là d'un constat qui s'oppose à la démarche macro-économique, où l'hypothèse centrale est celle d'une invariance des structures des agrégats.

Ce constat nous amène à expliquer l'objet de notre travail, qui n'est pas d'expliquer par exemple le niveau de l'emploi, l'état ou le fonctionnement du marché du travail en introduisant une modélisation ou des hypothèses, mais plutôt de soumettre un ensemble de données à un traitement statistique sans idées a priori, dans l'esprit de la statistique descriptive. Il ne s'agit pas alors d'analyser ces données en émettant des hypothèses privilégiant telle ou telle variable aux dépens des autres, mais au contraire, de rendre explicite la structure interne des données, laquelle pourra conduire à la formulation des hypothèses.

Pour ce faire, nous utiliserons le modèle Log-linéaire : il permet d'étudier une répartition d'individus ou une distribution de leurs caractéristiques dans un espace à plusieurs dimensions. En effet, l'observation à l'intérieur de chaque cellule d'un tableau des données multidimensionnel est décomposée en un produit d'effets s'appliquant à elle. Le modèle Log-linéaire renferme donc un caractère multiplicatif, et à notre sens est plus respectueux de la nature des grandeurs économiques, généralement des grandeurs multiplicatives.

L'utilisation de cet outil pour la modélisation des variables du marché du travail, notamment l'offre d'emploi salarié, permet de cerner l'imbrication et la multiplicité des facteurs qui l'influence, grâce à un ensemble de paramètres calculés simultanément dans une relation d'ensemble. C'est donc l'atout principal de ce modèle par rapport aux techniques descriptives multidimensionnelles traditionnelles, qui reposent sur l'appréciation visuelle des résultats graphiques, et ne disposent pas d'instrument de mesure des liaisons entre les variables et de leur effet sur la structure d'un tableau de contingence. Le modèle Log-linéaire est ainsi caractérisé par l'existence d'une relation mathématique pour la description de la structure d'un tableau de contingence, et permet des interprétations économiques et des prolongements analytiques qu'une représentation visuelle occulte en partie.

0.5 Plan du travail.

La présente thèse comporte sept chapitres se présentant sous une forme ternaire :

- les deux premiers chapitres font le tour des théories du marché du travail ;

- les trois chapitres suivants (chapitres 3, 4 et 5) sont axés sur l'aspect méthodologique, et traitent particulièrement le modèle Log-linéaire sous ses différentes approches ;
- les deux derniers chapitres (chapitres 6 et 7) constituent le champ d'application du modèle Log-linéaire aux données de l'offre d'emploi salarié en France. L'analyse des résultats fournit des éléments pouvant permettre de surmonter certaines limites relevées au plan théorique (chapitres 1 et 2).

Le premier chapitre comporte deux grandes parties :

- la première partie sert de cadre d'analyse théorique où l'on aborde la définition du marché (Cf. 1.1.1) et son application au facteur travail, en dépit de son caractère impropre (1.1.2). L'économie du travail est dominée par deux paradigmes rivaux, que FAVEREAU[9] caractérise, l'un, de «mythe du Marché Universel» l'autre, de «mythe du Système Inégalitaire», le premier renvoyant à la conception néoclassique du marché du travail (Cf.1.1.3.1) et le second aux théories de la segmentation.
- la seconde partie applique ce schéma théorique à la France, où l'on analyse l'évolution globale de son marché du travail de l'après-guerre à la crise. Deux grandes périodes sont retenues : celle de la forte croissance des années de l'après-guerre et celle qui intervient avec la crise. Une étude des grandes tendances du marché sera examinée sur ces deux périodes et ensuite, on observera si elles se reproduisent à l'échelon régional. Ce sera l'objet de la dynamique régionale de l'emploi (Cf. 1.2.2)

Le chapitre 2 met l'accent sur la localisation du marché du travail, c'est-à-dire ses rapports à l'espace. Un certain nombre de points y sont développés :

- le premier point met l'accent sur la définition de la notion de l'espace et la mise en relief des rapports à l'espace des agents économiques (Cf. 2.1). En effet, tout individu est localisé en un lieu, il en est de même des entreprises. Cette localisation est facteur des coûts et d'opportunités, ce qui constitue des contraintes devant peser indiscutablement sur le fonctionnement du marché du travail ;
- le deuxième point est consacré à l'intérêt de la prise en compte de la dimension spatiale dans le marché du travail, intérêt qui découle entre autres, de la diversité géographique des variables du marché du travail (emploi, chômage, salaire, mobilité des actifs, etc.). Il s'achève

sur l'approche spatiale du marché du travail dans l'optique du modèle concurrentiel, où l'espace intervient de manière exogène : c'est un cadre imposé où l'on confronte les outils du modèle traditionnel. Cette approche dissimule donc les structures du marché.

Le troisième point aborde les marchés locaux (bassins d'emploi, zones d'emploi, etc..) et les techniques permettant de limiter géographiquement ces marchés. Tous ces points mettent en lumière la nécessité d'intégrer au marché sa dimension spatiale, mais ils ne prennent aucunement en compte les effets de structure. Le modèle Log-linéaire, se présente comme un outil permettant de surmonter de tels problèmes.

Le chapitre 3 dresse un panorama des concepts de base servant de cadre définitionnel au modèle Log-linéaire. Les notions de tableaux de contingence multidimensionnels et de marges qui leur sont associées y sont examinées. Ces notions apparaîtront ultérieurement, soit explicitement, soit implicitement dans la formulation du modèle.

Le modèle Log-linéaire appartient à la famille exponentielle et nous présenterons quelques processus d'échantillonnage appartenant à cette famille (Cf. 3.3). L'hypothèse sous-jacente étant, que les données du tableau sont caractérisées par ces processus. Le principe général du maximum de vraisemblance (Cf. 3.4) et la présentation de quelques modèles linéaires ou multiplicatifs, afin de mieux situer le modèle Log-linéaire, seront également développés dans ce chapitre.

Quant au chapitre 4, il est plutôt centré sur la formulation proprement dite de ce modèle. Cette formulation découle du principe de minimisation de l'information discriminante, preuve que la théorie de l'information sous-tend le modèle Log-linéaire. Certes, cet aspect ne sert qu'à tester les liaisons du tableau (relations d'indépendance, de dépendance totale ou conditionnelle, etc.) et ne se prête pas aux interprétations économiques ou en termes de mesure de l'effet d'une variable quelconque sur la structure du tableau. Toutefois, les prolongements de ce modèle, inspirés des travaux récents de J.A. ZIGHERA(1985), seront développés (Cf. 4.3.2) afin de rendre interprétables ses paramètres. Ces travaux restent fondés sur la méthode du centrage des paramètres, et confèrent au modèle Log-linéaire le caractère d'un outil véritable de statistique.

Autant dans sa formulation de base que dans ses prolongements, le modèle Log-linéaire fait totalement abstraction de la dimension ordinale que peuvent receler les modalités des variables permettant de construire un tableau multidimensionnel. C'est ce point qui objective l'analyse entreprise au chapitre 5, où la prise en compte de l'ordre requiert une modification des paramètres du modèle Log-linéaire de base. Plusieurs perspectives sont concevables pour y parvenir. Elles vont de l'affectation de scores arbitraires aux modalités des variables dites ordinales (scores monotones croissants ou décroissants), à leur estimation par le modèle même. Dans ce dernier cas, on est conduit au modèle Log-multiplicatif ou Log-non linéaire (la décomposition du modèle n'est plus seulement linéaire des paramètres) qui présente un certain nombre de propriétés importantes. Ce dernier modèle a quelques similarités avec les méthodes d'analyses factorielles des correspondances, bien que le cadre conceptuel des deux approches ne soit pas identique. Néanmoins, «l'orthogonalisation» du modèle par le biais du centrage des paramètres nous donne une piste de recherche visant à attribuer au modèle Log-multiplicatif la «fonction illustrative» des méthodes factorielles (Cf. 5.2.3.3).

Le propos des chapitres 6 et 7 est d'analyser la structure de l'emploi salarié en France entre 1976 et 1981. Le chapitre 6 met l'accent sur les résultats statistiques et on y aborde successivement :

- la présentation des données qui sont celles des enquêtes Emploi de l'INSEE, appréhendées selon l'âge, le sexe, la profession et la région. Les nomenclatures des variables "profession" et "âge" sont trop importantes et leur réduction est effectuée ici de façon à permettre toute accession à l'information du tableau plus facile. On passe ainsi d'un tableau d'environ 260 000 cellules à un tableau de 5796 cellules pour une année donnée (cette réduction est à lier aux contraintes qu'implique la capacité de traitement des données) ;
- les différents résultats statistiques (Cf. 6.2) : choix du «meilleur» modèle, paramètres estimés du modèle Log-linéaire de base, des modèles d'association et également les paramètres centrés ;
- l'analyse des effets simples (Cf. 6.3), tant en niveau qu'en évolution. On notera surtout le fort pouvoir de discrimination de la variable "profession".

Le chapitre 7 porte sur l'analyse des effets de structure de l'offre d'emploi. Comme on le verra, l'offre d'emploi salarié par métiers est spatialement typée. Ce résultat reste lié en partie à sa composante structurelle que nous mettrons en lumière à travers un groupe de régions représentatives.

Le sexe est également une variable permettant de différencier l'emploi, et il apparaît une forte féminisation des emplois dans les métiers du tertiaire. Certes, ce résultat ne constitue pas en soi-même une nouveauté, mais ce qui est frappant, c'est de constater que la présentation paradigmatique mettant en parallèle la tertiarisation de l'économie et la féminisation des emplois ne saurait être valable au niveau régional. Il existe un paradoxe qui montre qu'à l'échelon régional, outre les effets de structure, interviennent aussi des effets spécifiquement régionaux. Ceux-ci justifient déjà à eux seuls la nécessité d'associer l'espace à l'étude du marché du travail.

L'âge a aussi une influence sur la structure de l'offre d'emploi, et les jeunes excellent soit dans les métiers de faible niveau de qualification (ce sont les métiers où la rotation est très forte et qui répondent aux contraintes d'adaptabilité du système productif), soit dans les régions où ils sont largement représentés (Nord de la France). Nous analyserons également le comportement des autres classes d'âge.

Il va de soi que toute méthode de traitement, aussi sophistiquée soit-elle ou aussi parfaitement adaptée à tel problème ne corrigera ni ne complétera les insuffisances des données dont nous disposons. Bien que les données des enquêtes Emploi soient passibles d'erreurs, l'outil d'analyse utilisé ici nous conduit à des résultats fort intéressants.

NOTES

- [1] MARCHAND O., MARC N.: La population active de 1875 à 1982 : les facteurs d'une forte croissance, *Eco. & Stat* n° 171/172, Nov./Déc., 1984, p. 5.
- [2] VINCENS J.: Les nouveaux aspects des problèmes de l'emploi, *Revue d'Economie Politique*, n° 1, Jan./Fev. 1979.
- [3] FITOUSSI J.P. : Emploi, structure et régulation, *Revue d'Economie Politique*, op. cité p. 83.
- [4] STOECKEL : Effet d'entreprise et structuration du marché du travail, *Annales de l'INSEE* n° 16/17 1974 p. 244.
- [5] DOERINGER P, PIORE M. : Internal labor market and manpower analysis, D.C. Heath, 1971.
- [6] CHAFFIOT E. : La contribution du CEE au développement de la recherche sur l'emploi, *Bulletin d'Information CEE*, n° 55, avril p. 5.
- [7] BLACKITT : Multivariate analysis : no so much technicas Weltauschaunng, *Biometrics*, n°30, p. 387.
- [8] DUBRULLE : L'analyse des problèmes de l'emploi à travers l'étude des entreprises, *Bulletin d'information CEE* n° 55 avril p. 7.
- [9] FAVEREAU O. : Evolution récente des modèles et des représentations théoriques du fonctionnement du marché du travail, *Colloque Structures du marché Travail et Politiques de l'Emploi*, 3-4 octobre, 1985 p. 2.

CHAPITRE I

**RAPPELS THEORIQUES ET EMPIRIQUES DE
L'EVOLUTION GLOBALE DU MARCHE DU TRAVAIL EN
FRANCE DEPUIS L'APRES-GUERRE JUSQU'A LA CRISE**

INTRODUCTION

Au cours du temps, les différentes écoles de l'histoire de la pensée économique ont donné de nombreuses acceptions au concept de marché, généralement liées aux fonctions qui lui sont attribuées. Au 18^e siècle, la notion de marché est associée à celle d'une loi naturelle, plus efficace que celle des hommes [1]. Pour MARX, le marché est la sphère apparente de l'échange des marchandises, propres au système capitaliste ; il correspond au mécanisme rationnel et optimal de la répartition des ressources [2]. Ces définitions, toutefois non exhaustives, accordent une place plus ou moins importante à l'échange.

Les analyses que nous ferons dans ce chapitre, tout comme dans les chapitres ultérieurs, porteront sur le facteur travail. Il importe alors d'indiquer en quoi la notion de marché est adaptée ou pas pour caractériser les échanges en travail. Au-delà, même si c'est par commodité, nous adopterons ce concept et retracerons l'évolution globale du marché du travail en France sur longue période, afin de montrer qu'elle est liée à l'évolution du système économique.

Ce chapitre comporte deux principaux points :

- le premier point sert de cadre d'analyse théorique. On aborde la définition du marché, voire son application au facteur travail. On constatera que l'objet de l'échange oppose les néoclassiques (le travail est considéré comme toute autre marchandise) et les Marxistes (ce qui est échangé, ce n'est pas le travail, mais une force de travail dont la mise en œuvre s'inscrit dans le temps au cours du procès de travail) ; quelques modèles de représentation du marché du travail seront également analysés. Globalement, on distingue deux types de modèles de marché du travail : le premier modèle est de type néoclassique et s'appuie sur le modèle concurrentiel, dont les limites, notamment en ce qui concerne certaines différences observées sur le marché du travail, sont justifiées par quelques amendements. Leur prolongement conduit à la définition des théories dites "locales" (théorie de la recherche d'emploi, théorie du capital humain, etc.). Le deuxième modèle est celui de la segmentation du marché du travail. Il s'oppose au précédent et explique les différenciations du marché de façon endogène du fait du rôle structurant de la demande du travail ou encore la propension par les entreprises à diviser le travail pour d'assouplir

les coûts de gestion du personnel, etc. On peut concevoir deux scénarii du processus de la segmentation : le scénario des marchés segmentés et le scénario du marché dualiste.

- Le deuxième point applique ce schéma théorique à la France. L'évolution globale de son marché du travail est cernée de l'après-guerre à la crise. On y retient deux grandes périodes pour l'analyse : la période allant de l'après-guerre au début des années soixante-dix, est caractérisée par la forte croissance. Les tendances lourdes marquant cette période sont celle d'une homogénéisation du salariat, à laquelle se superpose une tendance de son hétérogénéisation, la première l'emportant sur la deuxième. La deuxième période se situe dans la crise où la baisse de l'activité affecte durablement le marché du travail. Le fait marquant au-delà de la montée du chômage, est l'éclatement du collectif du travail.

Il est important de quitter ce schéma global pour observer si les tendances qui viennent d'être dégagées se produisent au niveau régional. Cette observation s'effectuera à travers la dynamique régionale de l'emploi, qu'on appréhendera sur trois principales phases de l'histoire économique récente : la première phase est celle des années cinquante, où il y a polarisation de l'emploi en Ile-de-France ; la deuxième phase intervient dans les années soixante avec la politique de décentralisation et la troisième phase se situe au-delà de 1975 avec la crise. Au total, les disparités importantes qui existaient à l'après-guerre se sont atténuées de telle sorte que durant les années soixante-dix voire les années quatre-vingts, il y a un rapprochement des profils régionaux en ce qui concerne l'emploi.

1.1 SCHEMAS D'ANALYSE THEORIQUE.

Dans le présent paragraphe, nous allons développer les éléments de la théorie liés au marché, voire au marché du travail. En effet, une fois saisie la notion du marché dans son acception générale (1.1.1), nous analyserons l'application de ce concept au facteur travail (1.1.2) avant de présenter quelques modèles de marché du travail (1.1.3).

1.1.1. Cadre définitionnel du concept de marché.

Ce cadre nous permettra de définir d'abord le marché comme un lieu abstrait, ensuite comme un lieu géographique. Nous terminerons sur la définition de quelques fonctions du marché.

1.1.1.1 Le marché, un lieu abstrait

Dans son acception générale, on peut dire que «le marché est un lieu, non pas géographique, mais abstrait, de rencontre et d'échanges d'informations et de biens, rencontres régies par des conventions explicites ou implicites qui dépendent du bien, des habitudes des échangistes, et d'un cadre institutionnel qui délimite au moins ce qui est interdit » [3]. Donc, cette définition s'appuie sur un système d'allocation basé sur l'échange.

Le marché est régi par des conventions, voire des lois ou des dispositions particulières de la part des agents. En ce sens, on peut dire que c'est une institution. On remarquera dans la précédente définition que le concept de marché fait abstraction de la dimension géographique. Cette vision est également partagée par l'économiste Jevons qui estime que «la localisation n'est pas indispensable ; les commerçants peuvent être disséminés...si ... ils se trouvent en étroite communication les uns avec les autres » [4]. L'auteur met plutôt en relief l'existence des corps de personnes «effectuant des transactions qui permettent de confronter l'offre et la demande d'un bien quelconque par des réseaux d'information, lesquels jouent ici un rôle central. Toutefois, ces définitions ignorent que l'organisation du marché, tout comme les échanges et les rencontres qui s'y effectuent, se font dans un cadre géographique se fait dans un cadre géographique défini. Il convient alors d'expliquer la notion du marché en tenant compte de sa dimension spatiale.

1.1.1.2 Le marché, un lieu géographique.

C'est dans l'espace que se réalisent les échanges des biens. Afin de mettre en relief la dimension géographique du marché, nous allons nous inspirer de la définition empruntée à Cournot : « ...les économistes entendent par marché non pas un lieu déterminé où se consomment les achats et les ventes, mais un territoire dont les parties sont unies par les rapports de libre commerce en sorte que les prix s'y nivellent avec facilité et promptitude » [5]. Il y a donc un rapport entre le marché et l'espace, en particulier le territoire. Mais cette association de la dimension géographique au marché va poser le problème des limites. Faut-il laisser ces limites s'établir par des dispositifs institutionnels fixant des zones à l'intérieur desquelles diverses conventions règlent les rapports d'échange ou faut-il laisser ces limites s'établir par les seules

forces du marché ? La seconde alternative est plus compatible avec la définition de Cournot où c'est le prix, résultant de la confrontation de l'offre et de la demande d'un bien, qui délimite le territoire. Derrière cette définition, se pose le problème de la stabilité des limites puisque toute évolution des prix les affectera. Nous reviendrons à ce problème de limites géographiques du marché quand nous aborderons l'étude de la dimension spatiale du marché du travail (Cf. 2.3.2). Il est intéressant, avant de continuer l'analyse, de préciser les fonctions du marché, lesquelles nous permettront, d'élucider plu loin, lorsque nous transposerons la notion d marché appliqué au facteur travail, quelques types de marchés proposés dans la littérature par les économistes du travail.

1.1.1.3 Les fonctions d'un marché

Un marché possède trois principales fonctions : la fonction d'allocation, la fonction d'information et la fonction de régulation.

Fonction d'allocation

C'est la fonction qui permet à chaque agent intervenant sur le marché de se procurer les biens dont il a besoin, au prix qui lui convient. Il s'agit alors d'une répartition inséparable du système du prix qui lui est associé.

Fonction d'information

Elle permet à chaque agent sur le marché d'acquérir les données indispensables à son calcul économique, c'est-à-dire le plan de production pour les producteurs ou le plan des dépenses pour les consommateurs. Le système de prix joue encore un rôle primordial dans cette fonction, car «sous une forme abrégée, par une sorte de symbole, l'essentiel de l'information est transmis aux seuls agents concernés »[6].

Fonction de régulation

La fonction de régulation assure le passage d'une structure de répartition à une autre, c'est-à-dire d'une structure d'allocation à une autre. Ce passage intervient lorsque les prix et les quantités des biens prévus ne se réalisent pas.

La présentation de ces éléments préliminaires n'est pas sans intérêt, puisque nous les retrouvons dans les développements ultérieurs. Revenant au concept de marché, il ressort que celui-ci peut être caractérisé comme le lieu de l'échange d'un bien quelconque entre offreurs et demandeurs. Nous verrons comment appliquer ce concept au facteur travail.

1.1.2 Le concept de marché appliqué au travail.

Dans son acception telle que nous venons de le montrer, le marché constitue l'une des modalités par lesquelles se font les échanges en biens. Ce paragraphe nous donne l'occasion de vérifier dans quelle mesure il peut représenter les échanges en travail. En d'autres termes, les conditions d'échanges en travail sont-elles identiques à celles des biens physiques ou revêtent-elles des formes particulières ? Le travail peut-il être considéré comme une marchandise ? Afin de répondre à de telles questions, il est souhaitable d'identifier au préalable ce qui est échangé sur le marché lors de la conclusion d'un contrat entre le salarié et l'employeur. C'est ce qui constitue le premier point que nous développerons, c'est-à-dire l'analyse du travail comme objet d'échange. Ensuite, nous ferons une synthèse sur les approches néoclassique et Marxiste pour ce qui concerne l'objet échangé.

1.1.2.1 Le travail : un objet d'échange ?

L'analyse du travail comme un objet d'échange donne lieu à la confrontation de deux approches qui s'opposent : l'approche néoclassique, et l'approche Marxiste.

1.1.2.1.1 L'approche néoclassique

Pour les économistes d'inspiration néoclassique, l'objet de l'échange sur le marché c'est le travail, considéré comme un bien, autant que toute autre marchandise. Le salarié offre en échange d'un salaire un certain temps de travail dont la quantité dépend de sa fonction de préférence. Il fait alors un arbitrage entre le temps de loisir et le gain (salaire) pour optimiser ses satisfactions. Le principe de cette optimisation est le même que celui appliqué aux marchandises (égalité des utilités marginales).

La demande de travail par l'entreprise est régie par le produit marginal en valeur. Elle est alors déterminée par les conditions techniques de son fonctionnement. Dans cette approche, l'une des hypothèses dominantes reste celle de l'homogénéité du travail, ce qui reflète la difficulté des néoclassiques à intégrer les éléments qualitatifs dans leurs analyses. En effet, le travail n'est pas homogène (il en est de même des satisfactions qu'il apporte) et certains approfondissements de la théorie néoclassique tentent de mettre en évidence son hétérogénéité (théorie du capital humain, notion de « bien composite » de K. LANCASTER [7]).

Cette conception néoclassique de l'objet échangé sur le marché, même avec des approfondissements, s'oppose à l'approche Marxiste.

1.1.2.1.2. L'approche marxiste

Chez les auteurs Marxistes, ce qui est au centre de l'échange sur le marché, c'est la force de travail conçue comme une capacité à mettre ultérieurement en œuvre, et non le travail comme dans la problématique néoclassique. Il s'agit de la cession temporaire d'une « potentialité dont la mise en œuvre, sans être totalement aléatoire, n'est que partiellement définie lors de l'échange » [8]. Il y a donc une distinction entre le travail et la force de travail et MARX le souligne d'ailleurs : « ce que l'ouvrier vend ce n'est pas directement son travail, mais sa force de travail dont il cède au capitaliste la disposition momentanée [9].

On comprend aisément les modalités de l'échange dans cette perspective. Certes, l'échange lors de la conclusion du contrat entre le salarié et l'employeur porte sur la force de travail, mais la mise en œuvre de celle-ci n'est pas instantanée, ce qui donne à l'échange son caractère non spécifié, la seule chose vendue ou achetée étant une potentialité à exploiter dans le procès du travail. Toutefois, il convient de noter que même si la force de travail constitue l'objet échangé, elle a comme support le travail car « le travail est source de la valeur et la base de l'échange dans tout système économique ». Chaque fois qu'il y a échange en valeur, il y a échange en travail.

Les auteurs Marxistes reconnaissent à la force de travail quelques caractères d'une marchandise, bien qu'il s'agisse d'une marchandise particulière. Cette thèse est réfutée par MALLET [10] qui estime que le salarié ne peut aliéner complètement sa force de travail lors de la conclusion d'un contrat. De plus, la force de travail n'est pas produite comme une marchandise : sa production se fait hors de la sphère économique.

L'approche Marxiste nous éclaire davantage par rapport à l'approche néoclassique, dans la mesure où la distinction travail-force de travail permet de mieux cerner l'objet de l'échange, bien que celle-ci ne soit pas « aliénable par un simple contrat ». Il importe maintenant de faire le bilan des deux approches pour mieux élucider la notion du marché appliquée au facteur travail.

1.1.2.2 Récapitulatif des deux approches et situation particulière du marché du travail.

Les approches que nous avons développées ont mis en lumière les éléments indiquant les particularités du marché lorsque l'on adapte au facteur travail. Si l'on revient une fois à la perspective marxiste, on constate que la force de travail n'est pas une marchandise quelconque : « la force de travail renferme..., du point de vue de la valeur, un élément moral et historique ; ce qui la distingue des autres marchandises » [11]. Alors, est-il justifié de parler du marché du travail ?

La relation entre le salarié et l'employeur est liée à l'incertitude. En effet, la mise en œuvre de la force de travail est à la fois individuelle et intégrée dans un procès de travail. Les différentes étapes de cette mise en œuvre peuvent faire l'objet des renégociations, car la relation conflictuelle entre les deux partenaires implique des rapports de force dont chacun cherche à modifier les règles pour en tirer des avantages. Ceci témoigne de la spécificité de l'échange en travail. Dans ce cas, le contrat sert plus de cadre à la mise en œuvre de la force de travail que celui de l'organisation de l'échange en lui-même. En ce sens, on peut dire que « le marché du travail n'est que le siège de la conclusion de (s) (ces) contrats, ensembles des premiers actes d'une relation de travail à travers laquelle se réalisera l'échange » [12]. Ces éléments permettent de particulariser le marché du travail.

D'autre part, lorsqu'un contrat est conclu, un bien physique change de main et passe du vendeur à l'acheteur qui en fait l'utilisation qu'il veut. Ceci n'est pas le cas de la force de travail pour les raisons déjà mentionnées plus haut. On peut également mettre en relief la particularité du marché du travail par rapport aux autres marchés à la lumière des mécanismes de fixation de prix. Le salaire n'est pas seulement déterminé par les forces du marché (confrontation de l'offre et la demande) ainsi que témoigne J. HICKS : « les employeurs seront réticents à augmenter le salaire à cause seulement de la rareté du travail : car le fait d'offrir des salaires plus élevés à certaines qualifications qui sont devenues rares aurait pour effet de bouleverser la grille des salaires. Ils sont réticents à baisser les salaires simplement parce qu'il y a du chômage, car s'ils faisaient ainsi, ils s'attireraient l'hostilité de ceux qu'ils continuent d'embaucher. » Le salaire perd alors le rôle d'un prix flexible.

Partant de ces éléments d'analyse, on constate que l'application du concept du marché au travail semble relativement impropre, son utilisation se justifiant essentiellement par sa commodité. C'est pour cela que certains auteurs préfèrent parler de système d'emploi, afin de traduire la complexité des influences qui s'exercent sur le marché. Bien que la terminologie du marché apparaisse gênante pour traduire les échanges en travail, nous allons néanmoins l'utiliser pour décrire quelques cas de marchés que nous propose la théorie économique.

1.1.3 Quelques modèles de marché du travail

L'économie du travail est traditionnellement partagée entre deux principaux modèles de marché : le modèle néoclassique et le modèle de segmentation. Ces deux types de modèles feront l'objet de l'analyse dans ce paragraphe.

1.1.3.1. Modèles néoclassiques du marché du travail.

Le paradigme néoclassique du marché de travail repose sur le modèle central qu'est le modèle concurrentiel, lequel a subi quelques transformations, du point de vue de l'offre du travail. Nous allons rappeler les hypothèses du modèle, leurs limites et les prolongements grâce aux " théories locales" néoclassiques.

1.1.3.1.1 Hypothèses du modèle concurrentiel.

Le modèle concurrentiel a été construit sur la base de quatre principales hypothèses :

- *l'homogénéité du travail* qui suppose qu'il n'est pas possible de distinguer les différents types de travail. Les qualifications sont substituables entre elles ;
- *l'information parfaite* : cette hypothèse porte sur les conditions d'accès à l'information par les agents sur le marché : chacun peut avoir facilement connaissance du prix du marché ;
- *la mobilité parfaite* sur le marché, qui met en évidence la libre circulation des agents sur le marché. Bien entendu, cela suppose un cadre spatial à l'intérieur duquel aucun obstacle n'entrave la propension à la mobilité.
- *l'atomicité des agents* : ceux-ci, pris individuellement, n'exercent aucun pouvoir sur le marché. Cette hypothèse est rejetée dans la formulation de certains types de marchés tels que le monopsonne.

Une fois présentées ces hypothèses, le modèle fonctionne par confrontation d'une offre globale et d'une demande globale, obtenues par sommation des offres et demandes individuelles de travail. Il s'établit un prix d'équilibre (salaire) qui apparaît comme la variable déterminante. Cette représentation présente des limites tant du point de vue des hypothèses que du point de vue du rôle du salaire.

1.1.3.1.2 Limites des hypothèses du modèle

L'hypothèse d'homogénéité du travail apparaît comme une limite du modèle car elle ne permet pas de relever les différentes caractéristiques du travail : « cette hypothèse d'homogénéité masque le fait que tout processus productif est organisé et que les emplois se situent les uns par rapport aux autres dans le cadre d'une division du travail qui n'est pas juxtaposition des tâches. Ceci masque fondamentalement l'hétérogénéité du travail, même si c'est au sein d'un ensemble ordonné » [13].

Quant à l'information, on ne peut prétendre l'acquérir facilement, du fait de la multiplication des lieux de contrats sur le marché du travail, qui contribue à complexifier l'information, la rendant foisonnante. Dans ces conditions, il est peu probable que chaque partenaire ait la possibilité de connaître de façon absolue toutes les données du marché.

Pour ce qui est de l'hypothèse de la mobilité, elle est dans l'univers néoclassique modulée uniquement par les facteurs économiques. Or d'autres facteurs interviennent et ont un effet sur la mobilité : les facteurs relationnels, les facteurs liés à l'environnement, etc.

D'autre part le rôle accordé au salaire comme variable d'ajustement est simplifié, puisque son évolution tend à dépendre de moins en moins des déséquilibres observés sur le marché du travail. Par ailleurs, il y a montée simultanée du chômage et des salaires en France à partir de 1968 [14]. Le salaire tend donc à s'ajuster au coût de la vie et non à la quantité du travail.

Ces limites ont montré qu'il y avait inadéquation entre le modèle de départ et la réalité des faits. Quelques apports ont permis d'améliorer le modèle concurrentiel.

1.1.3.1.3 Quelques apports et "théories locales" d'inspiration néoclassique

Par rapport au modèle concurrentiel, les néoclassiques ont innové en apportant des éléments nouveaux aux modèles. Nous analyserons ces apports et ensuite, nous rappellerons les théories qui en découlent.

Apports néoclassiques

Pour les néoclassiques, certaines différences qui apparaissent sur le marché peuvent se justifier par des comportements d'individus. Prenant la qualification comme critère de différenciation, ils estiment que le marché est constitué de sous-marchés, correspondant aux classes d'individus (chaque classe est associée à une qualification), substituables à court terme. Il y a alors remise en cause de l'hypothèse d'homogénéité du travail. D'autres apports portent sur le rôle du salaire, l'existence des déséquilibres et l'information.

- *le rôle du salaire* : il n'est plus unique et les différences structurelles du salaire sont liées aux différences de productivités individuelles et la diversité des préférences. On retrouve l'idée selon laquelle «les emplois les plus rebutants devaient être mieux rémunérés que les autres pour compenser leurs inconvénients » [15].

- *l'existence possible des déséquilibres* entre l'offre et la demande de travail : ces déséquilibres sont supposés temporaires et imputés au temps, indispensable à la formation pour acquérir une qualification.

- *l'existence possible d'imperfections quant à la diffusion de l'information*. Mais ces imperfections ne sont également que de nature temporaire.

Ces apports restent focalisés sur le salaire (la détermination et les différences de salaires). Ils ne suffisent pas pour expliquer d'autres différences intervenues sur le marché du travail, et qui sont justifiées grâce au développement des théories locales.

Les théories locales néoclassiques

Les "théories locales" d'inspiration néoclassique s'inscrivent dans le cadre du renouveau de la pensée néoclassique, qui est une tentative de répondre aux questions nouvelles apparues sur le marché du travail, et qui portent sur la fonction d'allocation du marché. Ce sont les théories de la recherche de l'emploi (job search theory), du capital humain et de la discrimination.

- *la théorie de la recherche de l'emploi* ou théorie de la quête d'emploi est une théorie qui s'inscrit dans l'ensemble des travaux sur les fondements microéconomiques de la macroéconomie. Elle est un élément de réponse aux problèmes qui entravent le bon fonctionnement du marché du travail, à savoir le caractère imparfait de l'information et l'incertitude qui en résulte sur la connaissance des emplois et des salaires. L'incertitude a ici un caractère stratégique, et pour M. AGLIETA «elle (cette incertitude) proviendrait de la méconnaissance du salaire qu'un individu muni d'une qualification déterminée peut espérer obtenir en changeant d'emploi, au regard de son salaire dans son occupation actuelle. Cette méconnaissance engendre une interaction entre les comportements stratégiques des individus et des firmes » [16]. Donc, le comportement stratégique de l'individu peut être de se consacrer du temps à la recherche d'un emploi plutôt que d'occuper un emploi dont il ne veut pas. On

comprend la tentative de cette théorie à intégrer dans le modèle du marché du travail, certains phénomènes tels que la coexistence du chômage et des offres d'emploi non satisfaites.

- *la théorie du capital humain* quant à elle s'efforce de justifier les disparités salariales sur le marché. Elle formule que les rémunérations sont liées au capital humain dans la mesure où la productivité de l'individu augmente avec la formation. On retrouve là l'hypothèse de la rémunération à la productivité marginale. Selon cette théorie, la formation est coûteuse, ce qui rend « l'offre des formés » rare. Cette rareté maintient la productivité marginale à un niveau suffisamment élevé pour expliquer les différences de rémunération. L'éducation est ainsi la seule source d'augmentation de la productivité. Cette théorie ignore d'autres potentialités telles que la capacité physique, les « effets de l'inné », du milieu de l'individu, etc., qui ont également une influence sur la productivité. Elle a pour avantage de faire progresser la théorie de l'offre de travail en rapprochant la formation et l'emploi par une logique du marché.

- *la théorie de la discrimination* essaie d'expliquer pourquoi certains travailleurs ayant la même productivité étaient différemment rémunérés d'une part, et pourquoi les caractéristiques de certains types de travail étaient liées à certaines caractéristiques non économiques des individus qui les occupaient, d'autre part [17]. G.S. BECKER [18] introduit la notion de « goût pour la discrimination » pour expliquer les pratiques discriminatoires très souvent liées à la race. Cette approche de la discrimination par le marché semble plus être un élément permettant « d'analyser les effets des facteurs sociaux sur le comportement économique en évitant, soit de les rejeter dans la catégorie des imperfections, soit d'abandonner trop vite la théorie néoclassique en se privant ainsi de son pouvoir analytique » [19].

Les théories locales reconnaissent donc l'existence des différences sur le marché du travail, mais elles tentent de les justifier du point de vue de l'individu (offre du travail). En cela, ces différences sont exogènes au marché et ne relèvent pas des mécanismes qui les induisent – ils sont inhérents au système économique – ce que R. BOYER identifie au rapport salarial, défini comme « un ensemble de conditions qui régissent l'usage et la reproduction de la force de travail, qu'il s'agisse de l'organisation du procès de travail, de la hiérarchie des qualifications, de la mobilité de la force de travail ou encore de la formation et de l'utilisation du revenu salarial ». On se situe là dans une perspective qui explique les différences du marché de façon endogène. C'est celle dans laquelle s'inscrivent également les théories de la segmentation du marché du travail.

On peut schématiser en concevant la segmentation comme «une bipartition mettant l'accent soit sur la demande (marchés interne et externe, soit sur l'offre (marchés primaire et secondaire), soit sur les relations entre l'offre et la demande (marchés locaux) » [23]. Sur la base de ce schéma, nous essaierons de proposer une typologie du processus de la segmentation.

1.1.3.2.2 Typologie du processus de la segmentation.

En s'inspirant des analyses que nous venons de faire d'une part, et des travaux de R. LOVERDRIGE et A. MOK [24] d'autre part, on peut distinguer deux scénarios du processus de segmentation. Le scénario des marchés segmentés et le scénario du marché dualiste.

Scénario des marchés segmentés

Le processus de segmentation se manifeste sur le marché par des stratifications que l'on peut appréhender à partir d'un découpage. Dans la perspective des marchés segmentés, il est possible d'examiner trois types de découpage [25] ; le premier est de nature professionnelle, le second est de nature géographique et le troisième est celui des marchés internes et externes.

Le découpage de nature professionnelle est fondé sur la théorie du capital humain. On suppose ici que la hiérarchie des professions repose sur le seul critère des qualifications acquises soit par la formation sur le tas, soit par l'éducation formelle.

Le découpage de géographique s'appuie sur la mobilité spatiale et c'est celle-ci qui délimite les marchés locaux. Il peut être effectué à partir des migrations alternantes qui constituent un des moyens de délimiter les bassins d'emploi, perçu comme des zones à l'intérieur desquelles la recherche d'un emploi et l'accès à un emploi se réalisent sans migration (au sens de changer de résidence).

Le dernier découpage est celui des marchés internes et externes. Cette division du marché résulte des procédures et règles administratives qui régissent le prix et l'allocation du travail. Il convient de noter que la notion de marché interne a été utilisée dans le cadre des théories de la segmentation en guise de réponse au problème posé par l'existence d'emplois cumulant durablement des avantages alors que d'autres cumulent durablement des désavantages. Il s'agit

1.1.3.2 Modèle de segmentation du marché du travail.

Ce paragraphe abordera deux principaux volets : des éléments généraux sur le processus de la segmentation et la typologie du processus de la segmentation.

1.1.3.2.1 Eléments généraux sur le processus de la segmentation

Les théories de la segmentation, contrairement à l'approche néoclassique du marché du travail, expliquent les différenciations observées sur le marché par le rôle fonctionnel dans le système d'utilisation de la main-d'œuvre. Ainsi, la segmentation doit être expliquée en termes de demande de travail par les entreprises et on retrouve ici le «paradigme de la demande», qui chez MERIAUX [20] constitue le point de convergence aux recherches françaises en économie du travail. Ce rôle dominant et structurant de la demande place désormais l'entreprise au centre de l'analyse et non plus l'individu tel que nous l'avons vu dans l'approche néoclassique : « le rôle de l'entreprise en tant que créatrice de l'emploi, en tant que source majeure de formation professionnelle et dispensatrice essentielle des rétributions dans notre société » [20] est d'ailleurs manifeste. L'entreprise demeure alors le lieu spécifique de l'allocation des ressources en main-d'œuvre et la segmentation apparaît comme un moyen par lequel elle acquiert face à l'incertitude des degrés de liberté pour modifier la structure des salaires, et créer une flexibilité dans l'utilisation de cette main-d'œuvre.

Le processus de la segmentation est inhérent au système économique et n'est pas un fait nouveau, ainsi que le montrent les propos que tenait cet agronome romain du premier siècle : « les terres malsaines, il vaut mieux les faire travailler par des mercenaires que par des esclaves » [22]. Cette propension à diviser, à segmenter le travail a également occupé les économistes institutionalistes tels que C. KERR qui, après la seconde guerre mondiale, insistent sur la nécessité pour les employeurs, face à des conditions incertaines du marché du produit, d'établir un noyau de travailleurs "permanents" auxquels devraient s'adjoindre, pour de courtes périodes, un certain nombre de travailleurs temporaires. L'inflation des monographies consacrées à la segmentation depuis les années soixante-dix est à lier à son ampleur grandissante qui résulte de l'incertitude qu'engendre la crise.

«d'une unité administrative, où la rémunération et l'allocation du travail sont gouvernées par un ensemble de procédures administratives » [26]. Ce marché s'oppose au marché externe du travail «propre à la théorie économique traditionnelle, sur lequel les décisions de rémunération, d'allocation et de formation sont contrôlées directement par des variables économiques »[27]. Il correspond donc au schéma du modèle concurrentiel (Cf. 1.1.3.1.1). Les deux marchés ne sont pas totalement indépendants : ils sont «[...] reliés entre eux, et les mouvements entre les deux ont lieu à certains niveaux de la classification des postes qui constituent des «portes d'entrée et sortie » du marché interne du travail »[28].

Afin de rendre l'analyse plus complète, nous présenterons également le scénario du marché dualiste.

Scénario du marché dualiste

C'est Doeringer et Piore qui ont été à l'origine du concept du marché dualiste ou marché dual. Ce concept est né aux Etats-Unis vers le milieu des années soixante, où l'on observait en permanence sur le marché du travail, des inégalités entre des groupes sociaux. La théorie du marché dual distingue deux secteurs : le marché primaire, aux emplois plus intéressants, c'est-à-dire offrant de meilleures possibilités de promotion et de meilleures conditions de travail ; ce sont les emplois les plus stables et les mieux rémunérés. A ce marché, on oppose le marché secondaire qui regroupe les emplois présentant les caractéristiques inverses : ce sont des emplois instables, mal rémunérés, mal protégés.

L'incertitude est le principal facteur d'explication du dualisme. En effet, celui-ci apparaît comme un des moyens de flexibiliser les coûts des entreprises dans une situation de variabilité de l'activité économique. Le marché a tendance à se fracturer en deux catégories d'emplois, chacune ayant son autonomie propre : «les divers segments de la société s'organisent autour de règles, de processus d'institutions différents, engendrant des systèmes, stimulants ou non, différents auxquels les individus réagissent»[29].

La théorie dualiste telle que nous venons de le voir révèle néanmoins quelques limites :

- elle repose sur l'importance accordée aux comportements. Certes, on comprend le souci des tenants du dualisme de se démarquer de la théorie traditionnelle en invoquant plusieurs types de comportements. Mais au-delà, ne reste-on pas toujours dans la logique d'une analyse du type néoclassique, dans la mesure où chaque type de comportement déterminant un segment, on peut appliquer à l'intérieur de celui-ci les outils du modèle concurrentiel, du fait de l'homogénéité des comportements à l'intérieur du segment ?

- il existe un problème d'identification des catégories. En effet, supposer que les femmes, les jeunes, voire les immigrés constituent un bloc marginalisé et homogène apparaît simpliste : « chaque catégorie (de ce bloc) remplit les fonctions spécifiques qui imposent de les distinguer entre elles ». De plus J. ROSE a montré que les jeunes à eux seuls ne forment pas une catégorie homogène.

Ce cadre d'analyse théorique est celui qui a longtemps dominé de nombreuses études liées au fonctionnement du marché du travail depuis l'après-guerre jusqu'à la période de crise dans les économies industrialisées. Chacun des modèles proposés tente de mieux représenter les faits réels inhérents au système économique auquel le marché du travail participe. Les différents modèles de marché du travail traduisent en quelque sorte les mutations du système économique. Celui-ci ayant subi de nombreuses transformations en France, nous analyserons sur longue période les nombreuses tendances qui ont marqué le marché du travail.

1.2 EVOLUTION DU MARCHE DU TRAVAIL EN FRANCE DE L'APRES-GUERRE A LA CRISE.

L'économie française a connu un ensemble de mutations depuis l'après-guerre, notamment avec la transformation de la structure des activités liée à l'exode rurale et agricole, à la restructuration de l'appareil productif et au développement des services, à la transformation des formes d'organisation du travail, etc. L'évolution du marché du travail ne saurait être dissociée de ces mutations de grande ampleur. Nous observerons ces transformations sur le plan global, puis nous analyserons leurs répercussions sur la dynamique spatiale de l'emploi.

1.2.1 Approche globale du marché du travail

Afin d'appréhender l'évolution globale du marché du travail en France, nous distinguerons deux grandes périodes : la période de forte croissance et la période de crise.

1.2.1.1 le marché du travail en France sur la période de la forte croissance

Ce sous-paragraphe nous permet de développer successivement les points suivants : l'évolution quantitative du marché du travail, les tendances d'homogénéisation du salariat et également les tendances d'hétérogénéisation du salariat.

1.2.1.1.1 L'évolution quantitative du marché du travail.

Comme il a été mentionné, les trente années de l'après-guerre sont caractérisées par la croissance exceptionnelle qu'a connue l'économie française sous l'effet de l'accélération du processus d'accumulation du capital. Les besoins en main-d'œuvre sont nombreux et entre 1950 et 1970, et l'emploi croît au rythme de 0,6 % par an [30]. Le plein emploi est assuré sans difficulté, et un haut niveau de l'emploi permet de satisfaire une demande d'emploi émanant de la population active disponible. L'ajustement est d'autant plus souple que cette main-d'œuvre peut prendre n'importe quel emploi qui lui est offert, quelle que soit sa nature, et ce dans la mesure où il exige un très faible niveau de qualification [31].

Parallèlement, le niveau du chômage est faible et oscille entre 1,5 % et 2 % de la population active, traduisant presque parfaitement les évolutions de l'emploi et de la population active, qui connaissent des taux de croissance relativement identiques.

La régulation de l'emploi dépend seulement d'une politique globale, fixant le nombre d'emplois offerts, et il n'y a pas de politique structurelle de l'emploi. L'accélération du processus d'accumulation du capital aura de nombreuses conséquences sur le salariat et sa structure. Au demeurant, deux tendances lourdes se manifestent et se superposent : une tendance à l'homogénéisation du salariat et une tendance à son hétérogénéisation croissante, la première l'emportant sur la deuxième [32].

1.2.1.1.2 Tendance d'homogénéisation du salariat

Les mutations induites par la croissance ont modelé le comportement des salariés dans le sens d'une homogénéisation. Ainsi que le note G.F. GERME [33], l'homogénéisation du salariat résulte de la conjonction de plusieurs facteurs :

- *les facteurs liés aux transformations du revenu salarial.* En effet, celui-ci devient pour un salarié, progressivement moins dépendant de sa contribution au travail collectif du fait de la sécurité sociale, l'indemnisation du chômage et la mensualisation des salaires. Dès lors, un salarié qui est malade ou en chômage court moins le risque de perdre son revenu ;

- *Les facteurs liés au caractère collectif du travail,* qui restent du point de vue institutionnel, le trait dominant de l'évolution du marché du travail depuis la guerre. Ces facteurs sont ceux touchant au lien entre employeurs et employés ; ce lien n'est plus tellement individuel puisqu'il est soumis au contrat collectif du fait du développement des conventions collectives au niveau des secteurs, des branches d'activités, voire des régions. Ainsi, les modalités de conclusion des contrats ou les caractéristiques et les problèmes de sa rupture ont tendance à s'unifier [34]. Ces facteurs ont eu pour effet de modifier les rapports de force sur le marché du travail, en renforçant la position des salariés, dans la mesure où ils limitent la liberté des employeurs quant à l'affectation des ressources en travail [35]. Cela a un impact sur le fonctionnement du marché du travail, ces facteurs ayant eu pour effet de favoriser les garanties d'emplois.

Toutefois, cette tendance à l'homogénéisation du salariat se superpose à celle de son hétérogénéisation croissante.

1.2.1.1.3 Tendance d'hétérogénéisation du salariat.

La période de croissance est également caractérisée par les mutations dans la structure des emplois, et qu'impliquent en partie les transformations de l'organisation du travail, qui se sont effectuées dans le sens d'une extension de la division Taylorienne du travail [36]. L'une des conséquences est manifestement l'accroissement de la part des ouvriers spécialisés dans l'ensemble des ouvriers.

Tableau 1 : Evolution taux du salariat en France entre 1954 et 1971

ANNEES	TAUX	ANNEES	TAUX
1954	64.7	1963	72.4
1955	65.6	1964	73.4
1956	66.4	1965	74.1
1957	67.7	1966	74.8
1958	68.6	1967	75.4
1959	68.8	1968	75.9
1960	69.5	1969	76.9
1961	70.3	1970	77.8
1962	71.2	1971	78.5

Source: GAMBIER D. : Le marché du travail, p. 47

D'autre part, il y a augmentation rapide de la population active salariée, et ce aux dépens de la population non agricole et agricole, d'une part, et sous l'effet de l'augmentation de la population active, d'autre part. Ainsi, le taux de salarisation augmente sensiblement passant de 64,5 % en 1954 à 78,5 % en 1971 (cf. tableau 1). Ces mutations structurelles tendent à différencier les diverses catégories du salariat : « si pour la société entière cette croissance de la part du salariat est un facteur d'homogénéisation, à l'inverse, elle est un facteur de différenciation interne au salariat dans la mesure où des ressources en main-d'œuvre nouvelle viennent constamment s'y intégrer de façon spécifique » [37]. C'est donc cette spécificité qui particularise chaque catégorie du salariat, renforçant alors son hétérogénéité.

A la régression de l'emploi agricole, on peut opposer une progression de l'emploi industriel et de l'emploi tertiaire (Cf. graphique1). C'est surtout l'emploi tertiaire qui évolue très favorablement puisque sa part dans l'emploi total passe du tiers au début des années cinquante à plus de 50 % au début des années soixante-dix. Certes, le tertiaire est caractérisé par son extrême diversité et on ne peut supposer une évolution régulière dans les différentes branches qui le composent. On observe également une baisse des travailleurs indépendants et une augmentation des emplois des cadres (18,6 % en 1975 contre 8,1 % en 1954).

Au début des années soixante-dix, la période de croissance exceptionnelle commence à marquer le pas, si bien que lorsque la crise intervient, les tendances enregistrées auparavant connaissent globalement un net infléchissement.

1.2.1.2 Le marché du travail en France sur la période de crise.

La crise amorcée au début des années soixante-dix se manifeste par un essoufflement de l'expansion économique commencée à l'après-guerre. Après avoir fait le point sur quelques grandes tendances sur la période, nous montrerons ensuite comment intervient l'éclatement du collectif du travail.

1.2.1.2.1 Les traits marquants de la période de crise.

La période de crise se caractérise par un ensemble de déséquilibres dont la manifestation la plus perceptible se situe au niveau de l'emploi qui ne croît plus qu'à un rythme inférieur à celui constaté sur la période de croissance : le rythme de croissance est de 0,27 % par an entre 1973 et 1980 contre 0,62 % entre 1950 et 1970 [38]. On peut lier cette baisse de l'emploi global au déclin de l'emploi industriel qui perd le rôle majeur joué jusqu'ici. L'industrie perd en moyenne 84000 emplois par an entre 1974 et 1979 alors qu'elle en créait en moyenne 64000 par an entre 1968 et 1973. Dans le même temps, l'emploi du tertiaire augmente, mais en dessous des proportions à compenser totalement la baisse de l'emploi industriel.

L'emploi crée n'arrive plus à être efficace et se posent à la fois le problème quantitatif du volume crée d'emploi et le problème qualitatif relatif à son contenu [39]. Par ailleurs, l'emploi devient instable – ou du moins une partie de l'emploi – du fait de la recherche d'une adaptation volume d'emploi – niveau de production qui passe entre autres, par une politique rationnelle de gestion de la main d'œuvre ou par des innovations technologiques. Tout ceci provoque le développement des formes nouvelles d'emploi et l'on assiste à l'éclatement de celui-ci.

1.2.1.2.2 Eclatement du collectif du travail

La situation d'incertitude née de la crise et sa persistance moduleront les caractéristiques comportementales des employeurs, lesquels manifestent une moindre propension à investir, faute d'une assurance sur les perspectives des débouchés à terme sur le marché. L'effet de ces comportements est patent sur le niveau d'emploi, car les entreprises n'embauchent le plus souvent que des effectifs nécessaires à l'augmentation prévisible de leur production.

Sur la période, la tendance est celle d'un éclatement des collectifs du travail, résultant du développement des formes nouvelles d'emploi telles que le travail à temps partiel, l'intérim, les contrats à durée déterminée, les stages de toute nature (travaux d'utilité collective, formation en alternance, etc.)

L'éclatement de l'emploi renforce l'hétérogénéité du salariat et il y a remise en cause de son unité observée sur la période de croissance. Les formes nouvelles d'emploi revêtent une importance particulière à l'occasion des créations d'emploi. Une étude CEREQ [40] montre que sur la période 1977-1980, pour un solde de 104 000 suppressions d'emploi, on enregistre en fait la disparition de 156000 emplois à durée déterminée et la création de 52000 postes à statut précaire. L'extension de ce type d'emploi implique qu'un grand nombre de personnes passent plus fréquemment par le marché du travail et «les séparations entre chômage, inactivité, emploi se sont brouillées ; à tel point qu'à l'idée de frontière, il vaut peut-être mieux préférer celle de halo, ou de frange » [41].

En somme, on constate qu'à la relative homogénéisation de la collectivité du travail observée sur la période de croissance, s'oppose son hétérogénéisation croissante en période de crise. La collectivité du travail est éclatée et c'est ainsi qu'au sein d'un même établissement, coexistent des groupes de personnels auxquels s'appliquent autant de statuts qu'il y a des sociétés différentes, juridiquement leurs employeurs » [42]. On a là l'illustration que l'employeur réel n'est pas souvent l'employeur légal, et les salariés sont quelques fois en présence d'un employeur écran (cas d'un sous-traitant par exemple).

Ces tendances rappelées, on est à même de se demander si elles connaissent de fortes modulations selon les régions, du fait même que les activités économiques, aussi bien celles qui sont en crise que celles qui s'adaptent, se trouvent diversement réparties sur le territoire (on comprend ici le lien implicite existant entre structure des activités et structure des emplois d'une région). Cette perspective nous interpelle à l'analyse, sur la période qui a été retenue jusqu'ici, des grands traits et la dynamique régionale de l'emploi.

1.2.2 Dynamique de l'emploi régional.

Les rappels esquissés précédemment ont montré que l'économie française s'est sensiblement modifiée depuis l'après-guerre, sous l'effet d'une croissance forte qui est allée de pair avec une transformation de l'espace dans le sens de ses différenciations, notamment en ce qui concerne les structures d'activités, qui ont une incidence sur la structure des emplois.

Au niveau d'une région, les structures d'activités sont constituées d'un ensemble de fonctions économiques qui révèlent sa position dans la division spatiale du travail [43], celle-ci ayant été plus ou moins induite par le redéploiement industriel des années soixante, qui avait pour but d'assurer un certain rééquilibrage de la localisation des activités sur le territoire.

L'analyse à mener ici sera axée sur deux points essentiels : le premier point abordera le processus de redéploiement géographique des activités afin de montrer comment il a façonné les espaces économiques français dans le sens d'une « disjonction fonctionnelle de la localisation des activités » à la quelle on peut associer la logique d'une spécialisation de la main-d'œuvre. Le deuxième point quant à lui abordera les grandes phases de la dynamique régionale de l'emploi.

1.2.2.1 Redéploiement géographique des activités et structuration des régions françaises.

Le redéploiement géographique des activités est un phénomène qui est resté à un niveau relativement faible dans les années cinquante avant de connaître un essor remarquable durant les années soixante, sous l'impulsion de la politique de décentralisation industrielle qui avait entre autres, pour objectif de rompre l'hypertrophie menaçante de l'Ile-de-France. Ce mouvement intervient à une période où il y a internationalisation du capital et intégration accrue

de la France dans le commerce mondial. D'ailleurs, l'un des soucis majeurs de la DATAR sur cette période a consisté à «intégrer l'espace français dans la socialisation internationale de la production, c'est-à-dire à articuler le niveau régional et le niveau international» [44]. Ce processus s'est déroulé en parfaite symbiose avec le déclin des formes traditionnelles de production agricole, et a totalement bouleversé le paysage économique des régions françaises.

D'autre part, le redéploiement géographique des activités s'opère dans le sens d'une fracturation des espaces, et ce afin de rendre adéquate toute forme de production aux caractéristiques de la main-d'œuvre dans une région. C'est ici qu'on observe l'une des conséquences les plus importantes de ce processus : à cette fracturation de l'espace, correspond une fracturation de la main-d'œuvre, ce qui traduit bien le fait que le redéploiement spatial des activités a eu pour effet de favoriser la division spatiale du travail.

Dans ce mouvement d'ensemble, l'Ile-de-France voit sa position se détériorer au profit des autres régions, surtout celles du Sud du Bassin Parisien et de l'Ouest. Si l'on observe son évolution dans le temps, on constate que le redéploiement se maintient dans la crise, tout en s'orientant vers l'ensemble des régions de l'Ouest et du Sud-Ouest. Il y a alors concordance de la géographie du redéploiement spatial des activités et les priorités définies dans le cadre de la décentralisation en faveur des régions de l'Ouest et du Sud-Ouest, ce qui témoigne que le redéploiement a largement bénéficié des actions d'aménagement du territoire. On peut ébaucher une typologie des régions françaises sur la base de leur spécialisation fonctionnelle dans l'industrie [45] :

- Région spécialisée dans la production des biens de production : **Ile-de-France.**
- Région spécialisée dans la production de biens de production et de consommation intermédiaire : **Provence Alpes Côte d'Azur.**
- Régions de structure productive moyenne : **Rhône Alpes, Alsace, Limousin, Sud et Ouest du Bassin Parisien, Sud-Ouest.**
- Régions spécialisées dans la consommation de bien de consommation intermédiaire : **Nord Pas de Calais, Lorraine, Auvergne, Languedoc Roussillon, Nord et Est du Bassin Parisien.**

- Régions spécialisées dans la production des biens de consommation finale : **Ouest, Franche-Comté.**

Cette structure sectorielle des régions contribuera le plus à la dynamique de l'emploi, celui-ci étant lié aux structures économiques dans une région. Certes, cette dynamique n'est pas uniforme à travers le temps et nous l'analyserons sur un certain nombre de phases de l'histoire économique récente.

1.2.2.2 Les grandes tendances de la dynamique régionale de l'emploi.

Sur la période 1954-1980, on enregistre une création nette de 2,4 millions d'emplois en France, dont près de la moitié en Ile-de-France (1,1 million d'emplois créés dans ce total en Ile-de-France sur la période). Alors que la plupart des régions développées bénéficient des créations d'emplois sur la période retenue, l'Ouest et le Sud-Ouest connaissent plutôt une stagnation ou une diminution de leur emploi (cf. tableau 2). L'évolution de l'emploi telle que nous révèle le tableau 2 est presque en adéquation avec la typologie établie précédemment et permet de caractériser les grandes régions suivant leur forme de production industrielle. En effet, les régions spécialisées dans la production des biens de production (Provence Alpes Côte d'Azur ; Ile-de-France) ont des taux de croissance de l'emploi nettement supérieurs à la moyenne nationale (12,6 %). Les régions de structure productive moyenne ont un taux de croissance de l'emploi moyen, tandis que les régions spécialisées dans la production de biens de consommation finale (Bretagne, Poitou-Charentes, etc.) ont des taux de croissance se situant nettement en dessous de la moyenne nationale. L'examen de ces taux ne doit pas nous conduire à conclure qu'il y a une certaine uniformité des tendances d'emploi sur la période 1954-1980. Ce serait ignorer les trois grandes phases qui ont marqué la dynamique régionale de l'emploi, notamment : la période 1954-1962, la période 1962-1975 et la période 1975-1980.

1.2.2.2.1 La période 1954-1962

Sur la période 1954-1962, l'emploi est très fortement concentré en Ile-de-France du fait de l'extrême polarisation des activités déjà rappelées au précédent paragraphe. Au cours de

cet intervalle de temps, c'est en moyenne 50000 emplois qui sont créés par an en Ile-de-France alors que les régions de l'Ouest, du Sud-Ouest, ainsi que le Limousin et l'Auvergne en perdent 45000 en moyenne [46]. Seules quelques régions de province, à savoir la Provence Alpes Côte d'Azur, la Haute-Normandie, la Picardie, la Lorraine et le Rhône Alpes enregistrent une progression de l'emploi. Il s'agit des régions de tradition industrielle. Cette période sera succédée par celle qui s'ouvre avec la décentralisation et le redéploiement des activités dans l'espace.

Tableau 2 : Evolution de l'emploi régional entre 1954 et 1980.

REGIONS	taux (en %)	REGIONS	Taux (en %)
PROVENCE-ALPES-CORSE	39.1	BOURGOGNE	5.0
ILE - DE -FRANCE	30.0	PAYS DE LA LOIRE	5.0
PICARDIE	20.6	BASSE NORMANDIE	2.8
HAUTE NORMANDIE	20.4	NORD-PAS-DE CALAIS	1.5
RHONE-ALPES	19.0	POITOU CHARENTES	-0.7
LANGUEDOC- ROUSSILLON	16.9	MIDI-PYRENNES	-2.1
FRANCHE COMTE	16.6	AQUITAINE	-3.4
CENTRE	16.5	AUVERGNE	-6.2
ALSACE	11.8	BRETAGNE	-6.3
CHAMPAGNE ARDENNE	11.6	LIMOUSIN	-17.9
LORRAINE	8.9	FRANCE	12.6

Source: Recensements de la population et estimation de la division des études régionales de l'INSEE

1.2.2.2.2 La période 1962-1975

La deuxième phase de l'évolution de l'emploi est celle au cours de laquelle s'effectue le redéploiement géographique des activités, marqué par une décentralisation des activités de

l'Ile-de-France au profit des régions du bassin parisien et de l'Ouest. Aussi, n'est-il pas surprenant d'observer une baisse des créations d'emploi en Ile-de-France, alors que dans les autres régions, l'emploi total augmente dans l'ensemble : le rythme de création d'emploi en Ile-de-France n'est plus que de 45000 par an ; ce rythme atteint 100000 en province. Il y a donc une redistribution spatiale de grande ampleur qui s'opère sur cette période, mais qui se fait de façon différentielle.

Les régions de la moitié Ouest et Sud-Ouest ont une situation beaucoup moins favorable avec une croissance pratiquement nulle, voire très nettement négative dans le cas du Limousin et le déclin de leur emploi qui prévalait sur la période 1954-1962 se trouve très largement endigué. Les créations d'emplois en Provence Alpes Côte d'Azur sont nettement supérieures à celles enregistrées en Ile-de-France qui se trouve légèrement dépassé par la Picardie et la Haute Normandie[47]. Elle est approchée par la Franche Comté, le Rhône Alpes et le Centre.

Au total, la période est marquée par un resserrement des rythmes de croissance de l'emploi entre l'Ile-de-France et les autres régions ; ce resserrement profite surtout aux principales régions de décentralisation (Ouest, Sud du Bassin Parisien), mais aussi à l'Alsace et à la Franche-Comté. Ce schéma global masque la nature des emplois transférés et ainsi que le montre une étude de AYDALOT [48]. Les emplois non qualifiés représentent 53 % des emplois transférés de l'Ile-de-France entre 1965 et 1975. En revanche le bassin Parisien enregistre 56 % des emplois non qualifiés sur la même période. On peut donc dire que si globalement il y a déclin de l'emploi en Ile-de-France, son amélioration structurelle est manifeste et elle se dote d'une structure de qualification au détriment du Bassin Parisien et de l'Ouest.

1.2.2.2.3 La période de crise (à partir de 1975)

La troisième phase de la dynamique régionale de l'emploi intervient avec la crise qui a sensiblement affecté l'activité économique et s'imprime sur le paysage géographique français avec les caractéristiques spatiales assez affirmées. Les régions les plus touchées par la rupture de la croissance sont les régions industrialisées, et plus exactement les régions couvrant le croissant Nord/Nord-Est, qui va de la Haute Normandie jusqu'au Rhône Alpes [49]. Le Nord-Pas-de-

Calais perd 19,2 % d'emplois industriels sur la période 1975-1982 ; la Lorraine subit des pertes d'emplois similaires au Nord-Pas-de-Calais. Quant à la Franche-Comté qui a enregistré une progression forte de l'emploi sur la période 1968-1975, ses effectifs salariés régressent au rythme de 1,6 % par an sur la période 1975-1982, et se situe dorénavant parmi les régions où la croissance de l'emploi est très faible.

L'emploi stagne dans les régions de forte croissance avant la crise comme l'Ile-de-France, la moitié Nord du Bassin Parisien et le Rhône-Alpes.

De 1979 à 1984, les tendances persistent et seules les régions du midi restent encore assez créatrices d'emplois, alors que la plupart des régions voient leurs effectifs salariés s'éroder. Il faut souligner que dans l'ensemble, l'Ouest et le sud-Ouest sont relativement peu affectés par la rupture de la croissance économique de 1974 et sur le plan spatial, on observe, compte tenu de tous les mouvements qui viennent d'être décrits, un resserrement progressif des rythmes de croissance de l'emploi, qui induit dans un certain sens un rapprochement des profils régionaux lorsque l'on examine l'emploi régional suivant les trois grands secteurs.

1.2.2.4 Rapprochement des profils régionaux

Sur la période 1954-1980, l'examen du tableau 3 permet de saisir la profonde transformation des structures françaises.

Tableau 3 : Evolution des secteurs d'activité entre 1954 et 1980.

ANNEES	1954	1975	1980
SECTEURS			
PRIMAIRE	27,0	9,5	8,8
SECONDAIRE	35,0	39,2	34,9
TERTIAIRE	38,0	51,3	56,3

Source : TURPIN : " Panorama économique des régions françaises".

A la lumière de ce tableau, on observe que l'emploi industriel dans les années quatre-vingts se situe au niveau de celui des années cinquante, tandis que l'emploi agricole régresse régulièrement au cours du temps, perdant près de 20 points en un quart de siècle. A l'inverse la part de l'emploi tertiaire ne cesse de croître et passe de 38 % en 1954 à 56,6 % en 1980. Ce mouvement s'accompagne d'une nette homogénéisation des structures régionales d'emploi ainsi qu'on peut le constater sur le graphique 1 (Annexe). L'observation de ce graphique indique de fortes disparités sectorielles en 1954 avec toutes les régions de l'Ouest et Sud-Ouest essentiellement agricoles. Sur la période 1954-1975, l'écart entre les deux régions extrêmes en termes d'actifs agricoles est passé de 49 % à 21 %. Pour ce qui est des effectifs industriels, l'écart est de 20 % en 1975 au lieu de 36 % en 1954. Quant au tertiaire, il a joué un rôle secondaire dans la réduction des disparités régionales, du fait de sa forte croissance peu différenciée sur l'ensemble du territoire. En somme, il s'agit d'un secteur dont la diffusion spatiale est presque homogène.

CONCLUSION

Ce chapitre nous aura permis de rappeler le cadre d'analyse théorique pour l'étude du marché du travail. D'autre part, l'examen des mutations en France au cours des trente dernières années de l'après-guerre montre l'existence d'une multiplicité des faits que la théorie du marché du travail doit interpréter et synthétiser dans un ensemble cohérent. Aussi, pouvons-nous dire que chaque modèle de marché du travail proposé est suscité par les faits.

Le caractère inopérateur du modèle traditionnel a conduit les néoclassiques au développement de quelques apports de façon à adapter la théorie aux faits observés. Ce développement reste limité puisqu'il met l'accent sur les comportements individuels à partir desquels sont justifiées les différences du marché du travail. Une telle conception est antinomique à celle que sous-entend la théorie de la segmentation du marché du travail qui ne pose le problème de l'emploi qu'en partant de l'entreprise, de son comportement et de la dynamique régionale de l'emploi ne constitue que l'une des données pour l'apprécier.

La dimension spatiale apparaît donc comme une variable importante pour caractériser les modalités d'affectation du facteur travail. Il n'est donc pas étonnant de voir les responsables politiques locaux du travail à l'heure où se posent avec acuité les problèmes de l'emploi, lesquels sont douloureusement plus ressentis à une échelle réduite. Ces recherches localisées qui ont permis de mieux appréhender le fonctionnement réel du marché du travail renforcent l'idée d'une nécessité de le spatialiser. Tel sera l'objet du chapitre suivant.

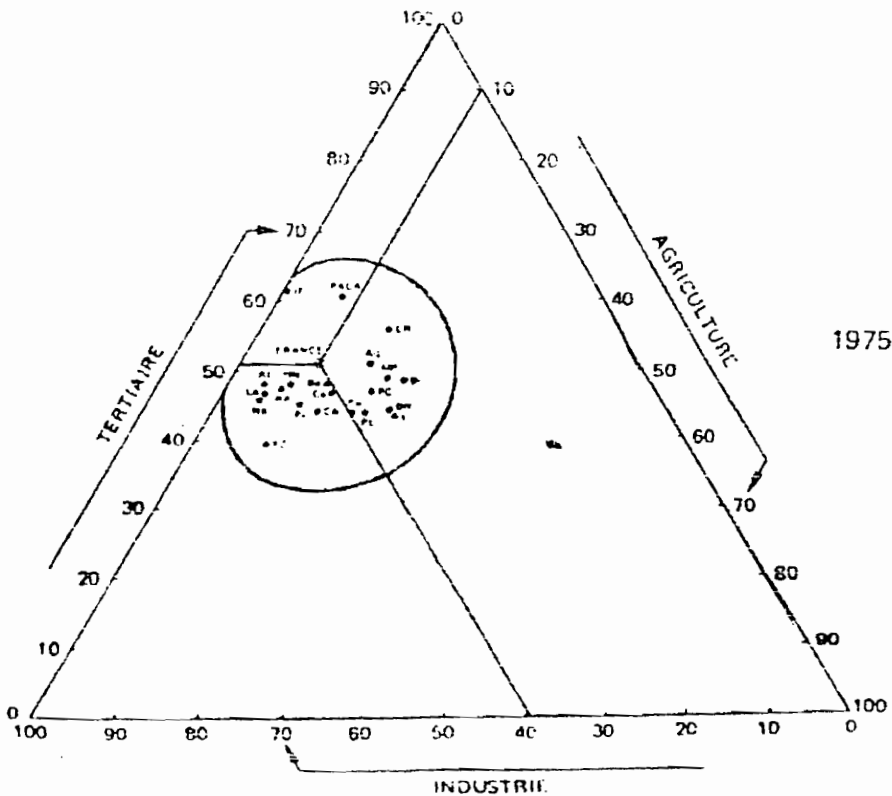
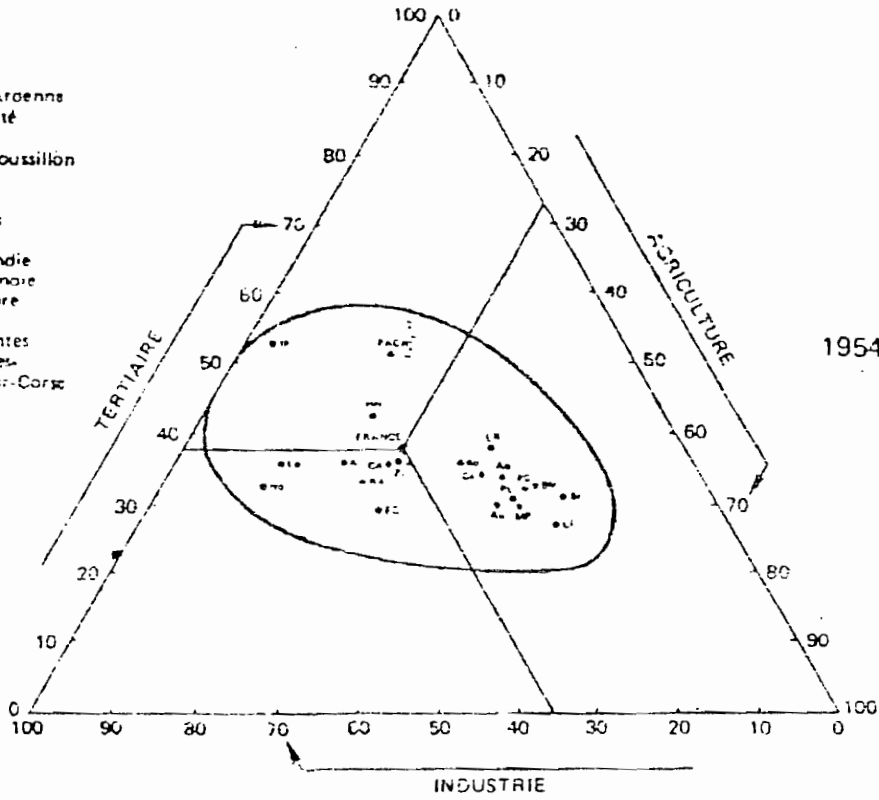
NOTES

- [1] Cette conception du marché est liée à l'idée de la «main invisible» développée par A. SMITH et reprise par M. BAUG (voir La pensée économique, *Economica*, 1978, p. 65).
- [2] Marx K. Salaires, prix et profits, Edition sociales, 1969.
- [3] MALLET L. : Le marché local du travail, Ed. CNRS, 1980, p. 9. On peut également donner au marché la définition empruntée à ABRAHAM-FROIS : « Le marché d'un bien peut être défini comme le lieu de la rencontre à un instant donné des désirs des consommateurs exprimés par leur demande, et de ceux des producteurs exprimés par leur offre » (voir Economie politique, *Economica*, p. 253).
- [4] JEVONS : Cité par MARSHALL : Principes d'économie politique, Paris, Gramma, 1971 T.2, p.3.
- [5] COURNOT : Cité par MARSHALL : Principes d'économies politique Paris, Gramma, 1971, T. 2, p.3.
- [6] MAYER F. : The use of knowledge in society, *American Economics Review*, 1945.
- [7] LANCASTER: Consumer demand. A new approach, N.Y. Columbia. Univ. Press 1971.
- [8] MARX K : op. cité p. 174.
- [9] GAMBIER D., VERNIERES M. : Le marché du travail p. 50.
- [10] MALLET L. : Op. cité p. 10.
- [11] MARX K : Le capital, T.1, p. 174.
- [12] MALLET L. : Op. cité p. 41.
- [13] HICKS : cité par LAL D. : Chômage et inflation salariale dans les économies industrielles, OCDE, 1977, p. 27.
- [14] GARNIER O. : Marchés internes et marchés externes du travail : leur rôle dans la formation des salaires, INSEE, 1985, p. 1.
- [15] VICENS J. : Les nouveaux aspects du problème de l'emploi, *Revue d'économie Politique* n° 1, 1979, p. 1
- [16] AGLIETTA M. : Panorama sur les théories de l'emploi, in *Revue Economique* n° 1, 78, p. 90.
- [17] EHRENBERG R.G, SMITH R.S. : Modern labor economics, Scott Foresman and Company, 1982.
- [18] BECKER G.S. : The economics of discrimination, Univ. Press, 2nd Ed., 1971.
- [19] ARROW K : Discrimination in labor market Princeton, Univ. Press, p. 4.
- [20] MERIAUX : Points de vue sur les recherches françaises en économie du travail, *Revue Economique* n°1, 78.
- [21] LOVERIDGE R., MOK A. : Théories de la segmentation du marché du travail, in *Marché du travail Européen*, CEE, Série Politique n° 42.
- [22] DUMRAD J. : Offre de travail et formes d'emploi, *Bulletin d'information du CEE*, n° 51, août 1981, p.1.
- [23] LOVERIDGE R., MOK A. : op. cité p. 22.
- [24] LOVERIDGE R., MOKA. : op. cité p. 27.
- [25] On peut constater que ce découpage met l'accent soit sur l'offre du travail, soit sur la demande du travail, soit sur la relation entre l'offre et la demande de travail.

- [26] DOERINGER P., PIORE M. : Internal labor market and manpower analysis, D.C. Health, p. 2.
- [27] GARNIER O. : op. cité p. 2.
- [28] GARNIER O. : op. cité p. 3.
- [29] AZOUVI A. : Théorie et pseudo-théorie : le dualisme du marché du travail, segmentation de l'emploi ou division du salariat ? Critiques d'économie Politique p. 7.
- [30] GIRAN J.P., GRANIER R. : Politique de l'emploi, *Economica*, p. 12.
- [31] VIMONT C. : L'avenir de l'emploi : l'illusoire, le possible, *Economica* 1891
- [32] Le trait dominant de la période de forte croissance est surtout une tendance à l'homogénéisation du salariat. Sur cette période on note notamment une mise en place des conventions collectives et d'autres législations qui tendent à unifier le salariat.
- [33] Germe G.F. : Instabilité, précarité et transformations de l'emploi, segmentation de l'emploi ou division du salariat ? CEP p. 80.
- [34] GAMBIER D. : Op. cité p. 34.
- [35] GAMBIER D. Op. cité p. 35
- [36] GERME G.F. : Op. cité, p. 81.
- [37] GERME G.F. : Op. cité p. 84
- [38] GIRAN J.P., GRANIER R. : Op. cité.
- [39] SALAIS R. : Les besoins d'emploi : contenu et problèmes posés par leur satisfaction, in *Emploi et Chômage*, op. cité p. 57.
- [40] AUDIER F. : La place de l'emploi précaire dans l'évolution de l'emploi (1977-1980) in *Formation et Emploi*, 1985, n° 10.
- [41] THELOT C. : Emploi et chômage : l'éclatement, *Eco & Stat* n° 193/194. Nov./Dec. 1986. p.3.
- [42] BROUDIC B., ESPINASSE J.M. : Les politiques de gestion de la main-d'œuvre, in *Travail et Emploi*, n° 6 Oct., 1980.
- [43] Voir VALEYRE A. : Dynamique régionale de l'emploi et division spatiale du travail
- [44] MACLOUF P., SARBIB J.L : Dualisation spatiale et dualisation sociale de la force de travail en période de crise. Deux scénarios possibles. *RERU* n°10, 1983 p. 123.
- [45] VALEYRE A. : Op. cité p. 10
- [46] VALEYRE A. : Op. cité p. 2.
- [47] TURPIN E. : Panorama économique des régions françaises, *Eco & Stat*, 1981
- [48] AYDALOT PH. : Mobilités des activités et de l'emploi. *RERU* n°3, 1978.
- [49] INSEE : Disparités et diversité des régions françaises à la veille du 9eme plan, INSEE.

EVOLUTION SECTORIELLE DE L'EMPLOI REGIONAL ENTRE 1954 ET 1975

- Al : Alsace
- Aq : Aquitaine
- Au : Auvergne
- Bc : Bourgogne
- Br : Bretagne
- Ce : Centre
- CA : Champagne-Ardenne
- FC : Franche-Comté
- IF : Ile de France
- LR : Languedoc-Roussillon
- Li : Limousin
- Lo : Lorraine
- MP : Midi-Pyrénées
- No : Nord
- BN : Basse-Normandie
- HN : Haute-Normandie
- PL : Pays de la Loire
- Pi : Picardie
- PC : Poitou-Charentes
- FACA : Provence-Alpes-Côte d'Azur-Corse
- RA : Rhône-Alpes



Graphique 1

CHAPITRE II

APPROCHE SPATIALISEE DU MARCHE DU TRAVAIL

INTRODUCTION

L'approche traditionnelle du marché du travail, telle que nous venons de le voir au chapitre 1, est par essence a-spatiale. Cependant, la nécessité d'y intégrer la dimension spatiale s'impose sous la pression des faits. En effet, l'étude régionale de quelques données socio-économiques révèle qu'un ensemble de résultats, obtenus à l'échelle nationale (à l'échelle globale), sont largement différenciés, une fois projetés à l'échelle régionale. C'est le cas notamment des résultats liés au chômage, à l'emploi, aux salaires, etc.

Par ailleurs, cette approche suppose, lorsque la dimension spatiale est prise en compte, une mobilité parfaite des travailleurs, afin de résorber les déséquilibres croissants qui se manifestent sur les différents marchés du travail régionaux. Cette hypothèse n'est pas confirmée par les faits, et l'on observe plutôt une certaine immobilité relative du facteur travail, c'est-à-dire la résistance à la mobilité géographique des travailleurs. Ce constat est le reflet implicite d'une différenciation spatiale du facteur travail ; les analyses globales ont tendance à masquer cette différence.

Ainsi, il semble judicieux de descendre à un niveau géographique réduit, où l'on peut envisager une analyse cohérente et pertinente du point de vue du facteur travail. C'est dans cette perspective que se situent les analyses locales du marché du travail, le local étant perçu comme «le niveau d'analyse qui permet de faire apparaître la diversité des comportements d'agents confrontés concrètement aux problèmes de l'emploi et du chômage, dans un environnement spécifique »[1]. A cette échelle réduite, on peut alors mieux identifier les agents intervenant sur le marché, et aussi appréhender leur stratégie. En somme, l'espace dans lequel se situe chaque agent va peser sur ses décisions et modeler son comportement. Ces éléments remettent en cause les analyses globales, auxquelles se greffent les analyses micro-économiques. Le haut comité de l'environnement dans son rapport sur l'amélioration de la qualité de l'emploi écrivait en 1978 : « à une approche globale qui laisserait échapper le caractère de recours de l'initiative locale, semblent devoir être préférées des expériences territoriales. Et cela pour plusieurs raisons : tout d'abord parce qu'elles permettront de coller au terrain et de rechercher, par petites régions, par bassins d'emploi, par pays, les besoins spécifiques d'activités qualitatives... »[2].

Le présent chapitre s'efforce donc de repérer les modalités d'une intégration de l'espace dans le marché du travail (les différentes transactions sur le marché relèvent d'agents localisés dans l'espace). Il s'organise autour de trois principaux points :

Le premier point essaie de cerner la notion de l'espace, et aussi d'établir comment les comportements des agents y sont différenciés (Cf. 2.1). Il apparaît que l'hétérogénéité de l'espace sert de base à des choix de localisation, qui seraient fonction soit de l'emploi disponible, soit des contraintes et d'opportunité. Autrement dit, les rapports des individus (2.1.2.1) et ceux des entreprises à l'espace (2.1.2.2) sont manifestes : l'individu est sensible à l'environnement dont l'espace est le support ; l'entreprise l'utilise comme modalité d'aspiration de la main-d'œuvre.

Le deuxième point est celui qui intègre l'espace dans le fonctionnement du marché du travail (cf. 2.2) Après avoir rappelé l'intérêt d'une telle intégration (2.2.1), l'analyse sera focalisée sur la prise en compte de la dimension spatiale dans le cadre du modèle concurrentiel (2.2.2), qui s'opère soit sur la base d'un critère (2.2.2.1), soit sur la base d'une modification de certaines hypothèses du modèle (2.2.2.2). Il reste néanmoins que l'espace ici apparaît de manière exogène, c'est-à-dire comme un cadre imposé à l'intérieur duquel fonctionne le marché néoclassique, ce qui constitue encore une limite de l'analyse.

Le troisième point met l'accent sur l'étude de quelques cas de marchés locaux (Cf. 2.3) : les bassins d'emploi (2.3.1.1), les zones d'emploi (2.3.1.2) et le cas des marchés locaux concrets (2.3.1.3). Au-delà, se pose le problème des limites du marché (2.3.2.), qu'on peut cerner à partir d'une démarche empirique ou statistique. L'existence des limites nous ramène au concept d'espace «géonimique», c'est-à-dire contenant un contenu et totalement indépendant des autres espaces. Le caractère a-structurel du marché limité géographiquement ne nous amènent-il pas à rechercher d'autres outils permettant de mieux élucider les caractéristiques du marché ? C'est ce point que nous aborderons au paragraphe 2.3.2.4.

2.1. CONCEPT DE L'ESPACE ET COMPORTEMENTS DES AGENTS FACE A L'ESPACE

L'objet du présent chapitre est de montrer l'intérêt qu'il y a de localiser un marché du travail, c'est-à-dire de mettre en relief ses rapports à l'espace. Dans les paragraphes qui suivent, il semble donc légitime, d'une part, de définir la notion de celui-ci (2.1.1.), d'autre

part, d'établir les rapports à l'espace des agents en présence sur le marché du travail (2.1.2). Ces rapports ont pour effet de fragmenter l'espace, mettant en évidence l'existence d'un facteur spatial dont on ne saurait ignorer dans l'étude des mécanismes du fonctionnement du marché du travail.

2.1.1 Notion d'espace

Le concept de l'espace, utilisé au sens large, recouvre un ensemble de définitions aux acceptions totalement différentes. On peut ainsi parler d'espace professionnel, d'espace visuel, d'espace géométrique, d'espace économique, etc.. Nous n'envisageons pas l'étude complète de cette notion, nous limitant simplement au concept permettant d'intégrer l'espace à l'étude du marché du travail. De ce fait, nous aborderons la notion d'espace économique, qui fait implicitement référence à un cadre géographique au-delà de ses caractéristiques, avant d'établir son lien avec l'aire du marché du travail.

2.1.1.1. L'espace économique

La conception de l'espace économique a déjà divisé un certain nombre de disciplines universitaires et comme le souligne J.R. BOUDEVILLE «l'étude théorique de l'espace économique est un maquis où les esprits les plus éclairés se perdent, et où s'affrontent les points de vue différents des géographes, des économistes, des mathématiciens... »[3]. Cette divergence de vue relève de la perception que chacune des disciplines fait de l'espace, et à cet égard J.R. BOUDEVILLE en distingue trois types :

- *l'espace géographique ou géonimique* qui définit un lieu où se situent et vivent les hommes. Il correspond à ce que F. PERROUX appelle « espace banal », cadre imposé, c'est-à-dire « un contenant, pour lequel le contenu est caractérisé sur le plan économique, ailleurs ». Il est évidemment repéré par des coordonnées géographiques (longitude, latitude, altitude).
- *l'espace mathématique ou « espace abstrait »* dans la terminologie de F. PERROUX. Il est constitué d'un système de relations abstraites, permettant de définir un objet, sans référence à une localisation géographique précise.

- *l'espace économique* qui fait la synthèse entre les deux types d'espaces précédents. Il est alors défini comme « une application de l'espace mathématique sur ou dans un espace géographique » [4], et traduirait ainsi les relations économiques entre les agents situés en un lieu.

Cette distinction éclaire davantage la démarche que nous entreprenons pour localiser le marché du travail. Celui-ci sera appréhendé comme espace géographique dans lequel il fonctionne si l'on se réfère à l'espace géonimique, ou simplement comme un espace défini géographiquement à partir de son seul fonctionnement si l'on fait référence cette fois-ci à l'espace abstrait. La première hypothèse suppose que le marché « contient un ensemble fermé d'agents et d'opérations, ce qui n'est pas le cas dans la seconde ». Il se pose implicitement un problème que nous aborderons ultérieurement : celui des limites géographiques du marché du travail. Autrement dit, celui-ci doit-il être supposé limité ou non ? Si oui, quels critères définir pour la limitation de l'aire du marché ? Peuvent-ils être pertinents ? Pour l'instant, on peut tout simplement constater que très généralement, l'aire du marché du travail est associée aux différents découpages administratifs, lesquels laissent apparaître l'autonomie de chacun des marchés non à travers son fonctionnement, mais par l'intermédiaire des conditions qui lui sont exogènes. C'est ainsi que, afin de prendre en compte certaines réalités locales, on a été amené à réduire très souvent l'étendue de l'espace, en passant de la nation à la région, puis de la région au département et à des « bassins » plus réduits sans aller au-delà des « espaces administratifs » [5]. Cet élément d'approche critique indique en même temps l'intérêt que pourrait susciter une approche spatiale du marché du travail à partir du concept de l'espace abstrait, puisque dans ce cas, il ne s'agit plus de considérer que les agents et les opérations sont contenus dans un contenant, mais de « reconnaître l'opposition entre espace cadre imposé à l'agent et l'espace économique engendré par lui » [6], voire les décisions qu'il prend. Les limites du marché dans ce cas sont de nature endogène, puisqu'elles sont fixées à partir du seul fonctionnement du marché.

Une fois cernée la notion d'espace économique, nous allons voir comment on peut progresser en élaborant d'autres formes d'espaces liées au marché du travail.

2.1.1.2 L'espace du marché du travail.

PERROUX [7] a pu définir quelques formes d'espace abstrait : l'espace comme « contenu du plan », l'espace comme « ensemble homogène » et l'espace comme « champ de

forces ». Il les caractérisera par rapport à la firme ; GAMBIER les reprendra et les situera dans la perspective du marché du travail. Il distingue :

- l'espace du marché du travail comme « contenu de plan » est « l'ensemble de relations parmi lesquelles figure une *fonction objectif*, celui-ci devant être atteint sur un certain espace géographique : c'est le cas de la réalisation d'une certaine croissance de l'emploi ou le maintien d'un certain minimum de chômage » [8].

- l'espace du marché du travail comme « ensemble homogène » est celui qui résulte de la prise en compte de certains critères permettant de décrire ainsi le marché. Il est alors possible de parler de marché des jeunes, de marché des femmes, de marché des ouvriers, de marché des cadres, etc.. L'un de ces critères au moins sous-tend la dimension géographique. D'ailleurs des études montrent que le marché du travail des ouvriers se réduit à une localité alors que celui des cadres est national. C'est ce qui ressort d'une étude de MACLOUF et J.L. SARBIB où l'on retrouve une opposition entre le couple stabilité sociale/ mobilité spatiale d'une part, et le couple mobilité spatiale/stabilité spatiale, d'autre part. Le premier couple est associé aux cadres et le second aux ouvriers [9]. Il s'agit là d'une dualisation de la force de travail qui se superpose à la dualisation spatiale de la dite force de travail.

- l'espace du marché du travail comme « champ de forces » définit un ensemble formel de relations liant des éléments entre eux. « Il est alors à la fois repéré par des pôles géographiquement fixés, entre lesquels existent des relations d'attraction ou de répulsion et un champ de forces centrifuges et centripètes » [10]. Dans ce cas, les discontinuités de l'espace ne sont pas exogènes, mais résultent des relations mêmes entre agents qui le déterminent. Cette conception de l'espace comme « champ de forces » correspond à l'espace polarisé ou encore champ de polarisation dans la terminologie de F. PERROUX.

LOIEZ LAURENT [11] montre que l'espace n'est pas homogène et qu'il est plutôt polarisé. Il comporte à cet égard trois niveaux de polarisation : la commune, l'arrondissement et la nation.

L'espace peut être également abordé dans la perspective marxiste. En effet pour MARX, l'espace géographique est le support des différenciations dans les coûts de reproduction de la force du travail, et dans le degré d'organisation des travailleurs. Dans ce cas, il est considéré comme un ensemble de lieux dans lesquels l'exploitation des travailleurs prendra des formes différentes. Ainsi, l'extension géographique des marchés a eu pour objectif, dans cette perspective, de combattre la baisse tendancielle du taux de profit, et cela

par une exploitation plus dure de nouvelles catégories d'individus, situées dans de nouvelles zones. Il s'agit là d'une perspective endogène inhérente à la logique d'accumulation intensive du capitalisme.

Ces quelques notions ne constituent que préalablement des éléments d'approche indispensables à la clarté des analyses que nous entreprendrons loin et qui mettent l'accent sur la nécessité de localiser le marché du travail. Avant d'y arriver, nous mettrons en relief les rapports des agents économiques à l'espace : les entreprises et les individus modifient par leur pratique l'espace qui leur induit à son tour des comportements. Cette interaction entre les agents et l'espace est au centre des mécanismes qui façonnent le marché du travail, et semble alors fondamentale d'être cernée lorsque l'on étudie les caractéristiques du marché.

2.1.2 Rapport des agents économiques a l'espace.

Dans le précédent paragraphe, nous avons défini l'espace et avons mis en relief le lien que celui-ci entretient avec le marché du travail. Ainsi, le concept du marché du travail est associé à celui d'une aire géographique et c'est surtout à ce niveau que sont pertinemment perçues les différentes relations entre agents économiques (individus et entreprises).

Sur le marché du travail, les conditions de la mise en œuvre de la force de travail sont déterminées par l'entreprise en fonction d'un contexte local, des contraintes et des opportunités d'environnement. L'utilisation de cette force par l'entreprise est réalisée dans un cadre géographique et celui-ci n'est pas sans intérêt pour l'individu. Ce cadre est évidemment en partie négocié lors du contrat initial, voire durant la vie du travail. Dans ce dernier cas, il peut entraîner une mobilité géographique de l'individu si l'avantage net qu'il en tire est positif.

Ces éléments d'approche indiquent implicitement que les agents économiques, notamment ceux qui interviennent le plus sur le marché du travail, ont des rapports spécifiques à l'espace, ce que nous analyserons par la suite en présentant d'abord les rapports de l'individu à l'espace, et ensuite ceux de l'entreprise à l'espace.

2.1.2.1 Les rapports de l'individu à l'espace.

L'individu a un certain type de rapport à l'espace qui se traduit par sa sensibilité vis-à-vis de l'environnement. Par ailleurs, même si l'espace lui impose des contraintes, il est aussi en mesure de le transformer. C'est donc ces points mettant l'accent sur la sensibilité de l'individu à l'environnement, et l'interaction entre l'individu et l'espace que nous aborderons ici.

2.1.2.1.1. L'individu est sensible à son environnement

Tout individu est situé en un lieu. Cette contrainte est très importante surtout lorsque ses choix dépendent de l'environnement immédiat, ou de son rapport de proximité avec les autres agents économiques. C'est ainsi qu'un individu qui recherche un emploi se situe dans un cadre géographique bien défini. Il y a donc un lien entre l'individu et l'espace de son action et ADAM SMITH le soulignait déjà il y a deux siècles, puisque disait-il « de tout ce qui peut se transporter, c'est l'homme qui est le plus difficile à déplacer » [12].

Dans l'analyse des rapports de l'individu à l'espace, ce qui est surtout important, c'est le degré de détermination que l'espace impose sur la destinée de l'individu, ce qui tend à prouver l'assertion suivante : « dis-moi où tu es né, je te dirai qui tu es ». Autrement dit, l'individu est le produit de son environnement, et il n'est donc pas surprenant que les comportements individuels varient d'un environnement à l'autre. D'ailleurs, la localisation de l'individu en un lieu reste tributaire d'un certain nombre de facteurs relativement liés à l'environnement : les facteurs affectifs (présence de la famille, d'amis, etc.), l'existence d'équipements collectifs (écoles, hôpitaux, équipements culturels, etc.), l'appréciation de l'environnement naturel ou climatique, l'existence des liens juridiques (propriété d'une maison ou d'un terrain), la présence d'équipements commerciaux (l'individu est aussi un consommateur) [13]. Au demeurant, ces facteurs diffèrent selon les lieux, du fait même du caractère hétérogène de l'environnement. L'espace apparaît dans ce cas comme le support de l'hétérogénéité de l'environnement. C'est un ensemble de localisations, c'est-à-dire d'environnements spécifiques.

2.1.2.1.2 L'interaction entre l'individu et l'espace.

Les rapports de l'individu à l'espace ont des caractères spécifiques que n'ont pas les marchandises et qui tiennent des capacités que l'homme possède en tout temps : « toute marchandise voit son identité transformée lorsqu'elle est placée dans certaines conditions physiques : chaleur, froid humidité... La différence avec l'homme est triple : sa tolérance à l'environnement physique est dans la plupart des cas très faible ; il est sensible à des caractères plus nombreux et plus complexes ; il est conscient de certaines influences et dispose de moyens pour les modifier »[13]. Il apparaît donc que l'individu et son environnement rétroagissent l'un sur l'autre, puisque au-delà des influences qu'a l'environnement sur l'individu, celui-ci « dispose de moyens pour le modifier ». Le fait d'être localisé à un endroit exige des contraintes spécifiques et l'individu ne peut s'en extraire qu'en changeant de cadre. Il est alors sensible à la localisation, capable de se déplacer et « les formes de la mobilité sociale ne sont pas indifférentes au lieu de la localisation initiale ».

Tous ces éléments d'analyse montrent en définitive que le comportement de l'individu est spatialement différencié. Toute tentative d'homogénéisation est un processus d'occultation du facteur spatial, qui devrait être intégré dans une analyse du marché du travail. Celui-ci est cohérent avec le système économique et induit un certain rapport des travailleurs à l'espace, du moins pour ce qui est du système de marché « capitaliste » : « la localisation, la fixité, la concentration des lieux de travail, la durée et les horaires de travail imposent un rapport à l'espace : le mode de vie compatible entraînera une limitation de la zone géographique habituellement parcourue par chaque individu (espace-étendue), une concentration des lieux de travail et de résidence (espace-environnement)»[15]. Les rapports de l'individu à l'espace tiennent donc au mode de socialisation cohérent avec le système économique. En fonction de sa durée maximale de déplacement quotidien lieu de résidence- lieu travail, l'individu tracera autour de son lieu de résidence, sa zone géographique d'emploi appelée zone de migration alternante, dont la limite dépend des infrastructures de communication ou des moyens de transport utilisés. Ainsi, l'individu dans son rapport à l'espace, vivra celui-ci comme un élément essentiel de sa vie quotidienne.

L'analyse que nous venons de mener peut être transportée à l'entreprise. En effet, celle-ci a un rapport à l'espace, mais qui diffère de celui de l'individu puisque ce dernier ne

sera pas sensible aux mêmes facteurs d'environnement que l'entreprise, celle-ci étant avant tout une organisation.

2.1.2.2 Les rapports de l'entreprise à l'espace

On peut aborder le rapport de l'entreprise à l'espace selon deux principes axes : d'une part, les stratégies de localisation, d'autre part, les relations qu'elle entretient dans son environnement, du fait du rôle central qu'elle joue dans les relations entre les autres agents économiques. Nous mettrons l'accent sur les stratégies de localisation, du moins dans leur version contemporaine, puisqu'elles relèvent des logiques qui se démarquent du schéma traditionnel. En effet, on observe de nos jours un éclatement spatial des fonctions de l'entreprise, ce qui reflète entre autres, l'aptitude à rechercher le travail le mieux adapté à ses besoins, et aussi le meilleur marché où qu'il se trouve. Cette faculté à choisir et à utiliser l'espace selon ses objectifs, est l'une des caractéristiques de l'entreprise, dotée d'une dimension importante. La petite entreprise voit ses possibilités d'action limitées dans l'aire où elle est implantée, et c'est cette aire qui la définit. Elle exprime alors les conditions qui y prévalent et n'en sera aucunement indépendante.

L'accent devant être mis, comme nous l'avons indiqué, sur les stratégies de localisation, nous rappellerons d'abord quelques éléments de la théorie traditionnelle en matière de localisation avant de montrer qu'ils sont inopérants dans le schéma contemporain, où la stratégie de localisation repose en grande partie sur le poids du facteur travail (HANNOUN et TEMPLE, 1975, AYDALOT 1978), l'entreprise cherchant à tirer parti de l'utilisation différenciée de l'espace en tenant compte des aptitudes et des qualifications de la main-d'œuvre située dans des cadres géographiques différents.

2.1.2.2.1 Éléments de la théorie traditionnelle de la localisation.

L'approche traditionnelle de la théorie de la localisation d'une entreprise en un lieu a longtemps mis l'accent sur le rôle joué par la distance et par conséquent les coûts de transport qu'il fallait minimiser. L'élément clé de cette théorie est la faible taille de l'entreprise,

ramenée implicitement à un point et disposant d'un marché lui-même circonscrit dans l'espace. On retrouve ici la notion de l'espace géonimique. Il y a donc une unité géographique de localisation de l'entreprise, ce qui suppose que toutes ses fonctions (direction générale, recherche et développement, production, etc..) s'accomplissent en un même lieu. Dans ces conditions, il est raisonnable que le meilleur choix de localisation soit celui qui permette de minimiser le coût total de production et de transport des produits finis et matières premières.

La stratégie de localisation peut être également abordée sous l'angle de la théorie des pôles de croissance. A la différence de la précédente qui concerne surtout les choix individuels de localisation, la théorie des pôles de croissance repose « sur une vision sociale de la croissance prenant en compte des facteurs économiques et non économiques » [16]. Elle met en relief la nécessité pour les entreprises de se concentrer dans de grandes agglomérations où elles tireraient parti des services qui y sont offerts et qui sont source d'économies externes ; ces agglomérations rassemblent également des entreprises complémentaires.

Dans tous les deux cas de figure l'entreprise a un rapport à l'espace. Celui-ci est facteur d'opportunités. Cependant ces théories sont incapables de décrire le comportement réel des grandes entreprises à l'heure actuelle, et qui se traduit par le recours à plusieurs espaces afin d'arriver à un certain nombre d'objectifs. Il y a alors remise en cause du modèle traditionnel.

2.1.2.2 Remise en cause de la théorie traditionnelle de localisation.

Depuis un certain temps, on assiste à la remise en cause de la théorie traditionnelle de la localisation. L'observation des faits et de nombreuses études (SALLEZ, 1983, AYDALOT, 1978) montrent que les grandes entreprises ne se sont pas étendues sur leur lieu d'implantation initiale, mais ont plutôt cherché à créer de nouvelles installations ailleurs. Elles se sont donc largement développées dans l'espace. Plusieurs facteurs ont favorisé ce développement, notamment les progrès techniques qui ont réduit les coûts de transport, d'une part, et facilité les liens entre les établissements séparés d'une entreprise, d'autre part. L'une des conséquences importantes de ce développement sera l'éclatement

spatial des fonctions de l'entreprise. Les travaux de PLANQUE [17] montrent une dualisation spatiale de ces fonctions : certaines fonctions tendent à se concentrer sur les principales agglomérations, alors que d'autres fuyant le surcoût de l'agglomération, s'implantent dans des zones traditionnellement peu urbanisées et peu industrialisées.

SALLEZ [18] nous cite l'exemple de RENAULT qui dispose des usines d'assemblage à Flins (Ile-de-France), à Sandouville (Normandie) et à Douai (Nord), mais approvisionnées en pièces détachées par des unités plus restreintes de Choisy-le-Roi et Rueil (Ile-de-France), Orléans et Dreux (Normandie). La plupart des unités n'approvisionnent pas un marché, mais plutôt une autre usine de la même entreprise. Les grandes entreprises créent ainsi des établissements fonctionnellement spécialisés, répartis dans des zones géographiques différentes. L'unité de l'entreprise, située en un lieu et accomplissant toutes ses fonctions en ce même lieu est donc remise en cause.

Derrière cette disjonction fonctionnelle de l'entreprise, se profile le processus de la division spatiale du travail qui intervient lorsque l'entreprise va au-delà de l'aire définie pour son activité voire son fonctionnement quotidiens. En ce sens, ce n'est plus l'aire géographique qui définit l'entreprise, mais celle-ci domine et définit l'espace au sein duquel elle agit. Nous mettrons en relief ce phénomène, en montrant comment l'entreprise fait recours à plusieurs types d'espace pour utiliser la force de travail dont elle a besoin.

2.1.2.2.3 Différenciation spatiale et caractéristiques de la main d'œuvre de l'entreprise

Les travaux de AYDALOT (1983) ont mis en évidence le poids du facteur travail dans les nouvelles stratégies de localisation des grandes entreprises, ce que reflète en partie l'éclatement spatial de leurs fonctions. L'auteur note que « des secteurs de plus en plus nombreux font éclater leurs activités de façon à aller à la rencontre d'une main-d'œuvre moins coûteuse » [19]. Dans la même perspective, ALAIN SALLEZ souligne que « les usines n'ont pas été implantées dans des zones spécifiques mais dans des zones où une main-d'œuvre bon marché et peu qualifiée préexistait » [20].

On constate donc que le coût et les caractéristiques de la main d'œuvre constituent l'un des enjeux majeurs pour l'entreprise. Cet enjeu est favorisé par les transformations importantes qu'ont connues les contraintes techniques ou financières de localisation, grâce aux progrès dans les communications où dans le transport, et qui se traduisent par une mobilité accrue du capital, correspondant à une homogénéisation de l'espace. A côté de cette homogénéisation du point de vue du capital, correspond une différenciation de l'espace plus marquée du point de vue du travail. L'entreprise utilisera en un lieu des types de main-d'œuvre les plus adéquats aux qualités exigées par le procès de travail. C'est ce qu'attestent d'ailleurs les études réalisées par MACLOUF et SARIB[21] où l'on assiste, dans le cas français, à une dualisation spatiale de la force de travail, c'est-à-dire l'émergence de deux types d'espace associés à deux types de main-d'œuvre :

- le premier type d'espace est formé des régions où sont localisées des activités de conceptualisation, les activités de tertiaire supérieur, etc..; la main-d'œuvre ici tend à être de haut niveau de qualification et jouit d'emplois stables.
- le deuxième type d'espace est constitué des régions où sont localisées des activités qui requièrent des techniques largement répandues, c'est-à-dire des activités de fabrication ou d'assemblage. La main-d'œuvre utilisée dans cet espace est « banale », c'est-à-dire qu'elle est interchangeable, peu qualifiée ; elle est abondante, bon marché et souvent caractérisée par un très faible degré de syndicalisation. Cette dualisation de l'espace français est à lier aux éléments de la théorie de la segmentation du marché du travail, ce qui nous conduit tout naturellement à la théorie dualiste du marché du travail.

En somme, bien que l'entreprise utilise dans l'espace les types de main d'œuvre dont elle a besoin, il est indispensable que le choix de chaque technique soit adapté à la main-d'œuvre recherchée. PH. AYDALOT, insistant sur ce point, indique que « chaque localisation exige une technologie correspondant aux caractères de sa force de travail ; chaque technologie exige une localisation offrant le travail qui lui correspond » [22]. Outre ces éléments d'analyse, il serait important de montrer le rôle de l'espace dans la mobilisation de la force de travail de l'entreprise.

2.1.2.2.4. Différenciation spatiale et modalité d'aspiration de la main-d'œuvre

Nous venons de voir que l'emploi est spatialement localisé selon les exigences techniques et l'adaptation de ces techniques à la main-d'œuvre. Nous allons nous affranchir de ces contraintes pour cerner comment l'entreprise recrute la main-d'œuvre dans l'espace.

Les modalités d'aspirations de la main-d'œuvre dans une aire seront liées aux caractéristiques de la main-d'œuvre recherchée et résultent d'un arbitrage quant à la « qualité » et à la « quantité » de cette main-d'œuvre. Ainsi, lorsque l'entreprise a besoin du personnel qualifié, elle se tourne vers les centres urbains pour leur recrutement. Dans le cas contraire, son champ de prédilection demeure les zones « périphériques ».

C'est ici qu'intervient l'importance quantitative des réserves réelles ou potentielles de la main-d'œuvre que l'entreprise peut utiliser localement. C'est également à ce stade que l'entreprise peut s'étaler dans l'espace, selon que les réserves potentielles existent ou pas. Autrement dit, si localement, les réserves réelles de l'entreprise s'épuisent, elle a la possibilité d'augmenter son aire de recrutement au cas où ses exigences en main-d'œuvre ne sont pas satisfaites. Le réservoir le plus utilisé est le stock de main d'œuvre rurale. Aussi voit-on le plus souvent les entreprises employant des ouvriers résidant dans un rayon de 30 voire 40 Km [23]. Bien que l'aire de recrutement s'étende dans l'espace, elle n'est pas homogène et dépend de la configuration de l'espace géographique ou de la répartition de l'habitat et du transport. Ce dernier point est important lorsque l'entreprise a sa propre politique de ramassage. L'étude effectuée par CAMPAGNA dans l'agglomération de Dunkerque [24] est à cet égard symptomatique. L'installation de Usinor à Dunkerque s'est opérée dans une situation de pénurie de main-d'œuvre. L'entreprise dans un premier temps a joué la carte de l'agglomération pour mobiliser sa main-d'œuvre, et ceci au prix d'une surenchère en matière salariale, dans la mesure où le marché du travail était cloisonné en de petits sous-marchés, contrôlés par les entreprises anciennes qui exerçaient une sorte de « paternalisme » sur la main-d'œuvre locale. A la fin des années soixante, une multitude de facteurs ont amené Usinor à se démarquer de l'agglomération à l'arrière pays et effectuer dorénavant son recrutement sur un rayon de 80 Km. Le système de ramassage par cars est ainsi passé de 46 % au début des années 70 à 70 % en 1975 [25]. Ce système présente un atout pour certaines

entreprises qui l'utilisent à leur avantage : il leur permet de faire éclater l'éparpillement de petits marchés dans l'espace. Il se pose le problème de la dimension de ces marchés.

Avant d'arriver au terme de ce paragraphe, il serait intéressant de procéder à une synthèse d'ensemble, qui permettra de mieux élucider les problèmes et les analyses à suivre ; nous mettrons beaucoup plus l'accent sur certains points que nous aurons rapidement survolé auparavant.

2.1.2.3 Quelques éléments de synthèse.

L'analyse des rapports des agents à l'espace montre que celui-ci influence de façon non moins significative leurs comportements. Un tel constat justifie au moins l'intérêt qu'il y a d'intégrer le facteur spatial dans l'étude du fonctionnement du marché du travail. Cependant, la prise en compte de cette dimension ne doit pas occulter le fait que les entreprises et les individus se situent dans des positions diverses face au marché du travail, ce que reflèterait d'ailleurs les comportements des uns et des autres, voire leur localisation dans l'espace.

Les individus sont sensibles à l'environnement. Ils peuvent se localiser en un lieu, compte tenu des contraintes et opportunités que leur procure ce lieu. Ils peuvent s'en extraire si les conditions leur sont défavorables. Néanmoins, on peut considérer dans l'ensemble qu'ils sont immobiles. Cet immobilisme relatif n'est pas à lier uniquement au coût de déplacement comme dans le cas des marchandises, mais à un ensemble de facteurs parmi lesquels, le coût de changement de lieu de résidence intervient significativement. Ce coût, dans ses différentes composantes, peut être plus important entre deux points d'un même espace qu'entre deux espaces assez éloignés. On comprend dès lors que l'environnement joue un rôle important dans la localisation de l'individu et il n'est pas surprenant que la reproduction de la force de travail soit plus spécifique et différenciée dans l'espace. Le coût de reproduction de la force de travail lui-même varie avec l'environnement (habitudes de consommation, structures démographiques, nature de l'appareil éducatif, infrastructures locales, etc.) .

De son côté, l'entreprise utilise l'espace et contribue à le structurer ; sa différenciation n'est en rien indépendante de la façon dont s'élaborent les politiques de main-d'œuvre.

Toutefois, il convient de remarquer que sur le plan spatial, ces politiques ne doivent pas être perçues uniquement du point de vue du coût différentiel de la main-d'œuvre, mais également s'interpréter à la lumière des conditions d'ensemble du procès de travail, notamment les potentialités d'utilisation des qualifications et des compétences des salariés. Autrement dit, l'entreprise est apte à structurer l'espace de façon à adapter le processus technique que requiert le procès de travail à la main-d'œuvre.

Il apparaît clairement qu'il y a décalage entre « espace de vie » (perçu comme un espace géonimique) et l'espace économique de l'entreprise. Si le découpage du premier semble ne poser aucun problème, il n'en va pas de même pour le second comme le montre l'éclatement fonctionnel de l'entreprise, qui reflète le fait que l'exécution d'un contrat de travail en un lieu peut faire intervenir un demandeur et un offreur de travail a priori très éloignés. Ceci est d'autant plus vrai que les stratégies des grandes entreprises se définissent plus à l'échelle nationale ou internationale, même si leur mise en œuvre ne peut s'opérer indépendamment de la prise en compte des facteurs locaux, notamment dans leur relation à la main-d'œuvre. C'est pourquoi dans la relation d'échange entre le salarié et l'employeur, outre le salaire, il est nécessaire de prendre en compte les phénomènes de proximité, lesquels interviennent en partie. Dans ce cas, on peut conclure que « l'intégration de l'espace dans le fonctionnement du marché du travail, plus qu'une question de frontières, est celle de l'articulation des relations entre agents, prenant en compte leur situation géographique, pour expliquer ce fonctionnement » [26]. Certes nous avons occulté ici les relations entre entreprises, lesquelles portent sur le processus de répartition de la main-d'œuvre et participent de ce fonctionnement du marché. Ces relations vont de l'entente (pour freiner la revendication des salariés et éviter une surenchère salariale ou encore constituer des bassins d'emploi avec relativement peu d'interférences) à la complémentarité (certaines entreprises embauchent pour des raisons variables des types spécifiques de main-d'œuvre et d'autres pas) voire de domination (en général les entreprises de grande taille ont une situation privilégiée sur le marché). Ces relations s'articulent également dans l'espace.

Outre les éléments que nous avons recensés et qui déterminent les rapports des agents à l'espace, il reste également que les mutations du système productif ont un rôle à jouer, puisqu'elles permettent de remodeler le paysage économique des régions. En somme, on peut

dire que les « rapports de l'agent à l'espace tiennent d'abord de la conception que celui-ci fait de l'espace économique en fonction de sa taille, des moyens techniques et financières qu'il met en œuvre, de sa capacité à le transformer [...] puis aux caractéristiques du mode de production au moment où s'exprime ce rapport en fonction duquel ses attitudes ne sont que le reflet d'une adaptation aux mutations du système productif » [27].

Ce paragraphe nous apparaît assez riche sur le plan analytique, dans la mesure où il remet en cause certains aspects de la théorie néoclassique du marché du travail, et aussi nous esquisse d'autres voies à suivre pour la perception judicieuse de son fonctionnement, celles qui prennent en compte les structures et les interrelations spatiales entre les différents marchés. Nous y reviendrons. On peut au vu des analyses que nous venons de faire, tirer deux conclusions partielles :

- le salaire est un élément déterminant de différenciation des coûts de production dans l'espace, puisqu'il dépend du coût de reproduction de la force de travail, lequel est largement défini par les conditions locales. Il s'agit là d'une vision opposée à celle de la théorie néoclassique qui soutient qu'il y a toujours neutralité du facteur salarial en tout point, toute différence de salaire ne matérialisant qu'un écart de productivité. La relation salaire-productivité est donc remise en cause.
- la structure du travail est prise en compte : les entreprises recherchent pour leur localisation, une structure bien définie de qualifications et d'expériences. Le facteur travail n'est plus considéré comme homogène.

Le facteur spatial n'est donc pas neutre, d'où l'intérêt qu'il y a de l'intégrer dans l'étude du fonctionnement du marché du travail. Ce facteur n'apparaît pas seulement dans la diversité des comportements des agents économiques en des lieux géographiques. On peut également l'appréhender à travers un ensemble d'autres facteurs que nous aborderons dans le paragraphe suivant.

2.2 INTEGRATION DE L'ESPACE DANS LE FONCTIONNEMENT DU MARCHE DU TRAVAIL

Les rapports des agents à l'espace ont mis en lumière des caractéristiques permettant d'observer des différences spatiales tant du point de vue du facteur travail (les individus) que des entreprises. On ne peut alors se prévaloir de gommer le facteur spatial dans les mécanismes de fonctionnement du marché du travail ; d'ailleurs une telle opération, conduit au constat d'un décalage entre les faits observés et le cadre analytique préétabli. Ce premier point sera complété ici par d'autres approches qui tentent de prendre en compte la dimension spatiale en vue d'améliorer l'efficacité des problèmes qu'elles posent. Cette section s'organise autour de deux paragraphes ; le premier retrace l'intérêt d'une localisation du marché du travail (2.2.1) alors que le second aborde la prise en compte de l'espace dans la perspective du marché concurrentiel (2.2.2)

2.2.1 L'intérêt d'une localisation du marché.

Outre les rapports que les agents entretiennent en des lieux géographiques un ensemble de phénomènes liés à l'espace apparaissent, suscitant l'intérêt de localiser le marché du travail.

Ces phénomènes résultent de la diversité spatiale des variables du marché du travail, c'est-à-dire l'emploi, le chômage, les migrations, etc.. On pourra également citer les phénomènes liés aux approches macro-économiques qui révèlent souvent l'extrême nécessité de descendre à un niveau géographique fin pour cerner les contraintes ou aspirations sans lesquelles les mécanismes globaux sont voués à l'échec. C'est le cas des modèles de développement régional qui ont prévalu dans les pays industrialisés durant la période de forte croissance. C'est aussi le cas des politiques de l'emploi qui sur cette période, se sont caractérisées par leur approche très globalisante. Nous montrerons l'insuffisance de ces deux approches et leur éventuelle amélioration par la prise en compte des spécificités régionales, lesquelles traduisent implicitement l'existence d'un facteur spatial qui, dans l'approche du fonctionnement du marché du travail ne peut être ignoré. Nous analyserons successivement la diversité géographique des variables du marché du

travail, le renouveau du paradigme du développement régional ensuite, et enfin, la nécessité d'intégrer les stratégies locales dans les politiques d'emploi.

2.2.1.1 Diversité géographique des variables du marché du travail.

L'étude statistique d'un certain nombre de variables du marché du travail nous conduit à observer des différences géographiques, notamment en ce qui concerne l'emploi, le chômage, l'activité, le salaire et la mobilité. Elles apparaissent tant en niveau qu'en évolution lorsque l'on examine quelques études régionales [29] portant sur ces variables.

Pour ce qui est de l'emploi, des différences régionales s'observent, que l'on retienne l'emploi total, industriel ou tertiaire. En 1985, plus d'une personne sur cinq occupe un emploi en Ile-de-France, une personne sur dix en Rhône-Alpes, environ une sur quinze en Provence-Alpes-Côte-d'Azur ou dans le Nord Pas-de-Calais [29]. A l'inverse, les régions suivantes ont une part faible dans l'emploi total : Franche-Comté (1,9 %), Limousin (1,3 %) et Corse (0,4 %). En évolution, l'emploi total baisse en Ile-de-France (24,2 % en 1984 contre 25,4 % en 1968) ; cette baisse est également manifeste dans le Nord-Pas-de-Calais et en Lorraine, alors que dans les régions de l'Ouest (Bretagne, Pays de la Loire), l'emploi total a le plus augmenté (2 % en moyenne par an après 1974). Ces différences se retrouvent dans l'emploi industriel où il y a un déclin relatif entre 1975 et 1981. Sur cette période, les taux de croissance annuels moyens varient de 3,1 % dans le Nord-Pas-de-Calais, - 2,9 % en Lorraine à -0,3 % en Alsace et 0,1 % en Bretagne pour ne moyenne nationale de 1,55 % [30]. Ainsi, quel que soit le critère retenu, on constate que l'emploi est fortement différencié dans les régions.

Quant au chômage, il est également diversifié sur le plan régional. Pour un niveau moyen de 4,3 % en 1975, les taux les plus forts sont enregistrés en Languedoc-Roussillon (5,9 %) ; la Bourgogne (3,2 %) le Centre (3,3 %) et l'Alsace (3,6 %) sont parmi les régions ayant un taux de chômage faible. En 1982, le chômage a plus que doublé sur le plan national (9 %) par rapport à 1975. Cependant, les régions les plus touchées par le chômage en 1975 le sont encore en 1982 ; celles qui étaient favorisées de ce point de vue le sont restées ; l'Alsace a un taux de chômage de 7 % et le Centre 7,7 %.

En ce qui concerne les taux d'activité féminins, les disparités sont aussi sensibles d'une région à l'autre. En 1975, pour une moyenne nationale de 39 %, les écarts vont de 49,6 % en Ile-de-France à 26,7 % en Languedoc-Roussillon. Les régions du Bassin Parisien ont un taux qui se situe à la moyenne nationale. Dans les régions du Sud, le faible taux d'activité qu'on y enregistre traduit les adaptations spécifiques qui se portent sur le fonctionnement du marché du travail.

Les salaires sont caractérisés par une forte dispersion spatiale. En 1975, le salaire moyen en Ile-de-France était de 51 % plus élevé que celui des deux régions les plus défavorisées : Poitou-Charentes et Limousin. Ce salaire moyen varie de 36948 Francs en Ile-de-France à 24447 Francs en Poitou-Charentes. Toutefois, les écarts entre les régions de province sont moins importants, puisque la région Haute-Normandie qui suit l'Ile-de-France ne devance le Poitou-Charentes que de 1,9 points. La plupart des autres régions se situent dans une fourchette comprise entre 26000 et 27000 de salaire moyen annuel.

Enfin, d'une région à l'autre, les disparités des comportements migratoires sont perceptibles. Si l'on examine la proportion d'actifs qui ont changé de région entre 1975 et 1982, on enregistre des variations qui passent de 3,9% dans le Languedoc-Roussillon (ces deux dernières régions étant parmi celles où les fourchettes allant de 5,1% en Languedoc-Roussillon) à -4,5% en Lorraine sur la même période.

Toutes ces différences témoignent de l'existence d'un facteur spatial et par conséquent, la nécessité de rendre compte des rapports entre l'espace et le marché du travail quand on étudie le fonctionnement de celui-ci. Cette nécessité apparaîtra également quand nous analyserons le nouveau paradigme du développement régional.

2.2.1.2 Renouveau du paradigme du développement régional.

En matière d'économie et de planification spatiale, les années de forte croissance, ont été marquées dans les pays industrialisés par la doctrine de développement « par en haut » [31] qui « fait confiance aux processus spontanés de l'économie pour entraîner à terme une

généralisation du développement ». Cette doctrine repose sur la théorie des pôles de croissance et suppose que les pôles les mieux lotis devraient générer des mécanismes susceptibles d'enclencher un processus de développement auto-entretenu dans les régions à développer. Cependant, l'observation des faits montre l'inadéquation de cette théorie à la réalité. Au contraire, il s'est produit, notamment en France, un développement sempiternellement inégal ainsi que l'attestent les travaux de PLANQUE [32], de AYDALOT [33] voire de SALLEZ [34]). Cette inégalité oppose les régions riches, cumulant un haut niveau de qualification, les capacités d'innovation, les économies externes, etc... à des « périphéries prolétarisées dont la croissance de l'emploi déqualifié serait alimentée par la seule décongestion des pôles ». Ces éléments ne sont pas nouveaux puisque nous les avons déjà implicitement soulignés quand nous avons analysé le rapport de l'entreprise à l'espace. La remise en cause d'une telle approche a fait renaître un autre courant de pensée en matière de développement dans l'espace : le développement « par en bas » ou « développement endogène ». Certains parleront de « mobilisation des potentialités locales ». Il s'agit d'une démarche centrée sur l'idée d'une économie organisée pour et par les communautés localisées : elle témoigne également d'une nécessité de rendre le développement adéquat aux réalités locales. Dans cette perspective, il devient indispensable de définir des unités spatiales de base. On parlera alors de « micro-régions », de « terroirs », de « sociétés locales », d'écosystèmes ». Ces différents vocables renvoient à l'espace pris comme le « territoire de vie d'un groupe doté d'une sorte de droit au développement local » [35] et non comme une distance entre des lieux. Or nous l'avons vu, les comportements des groupes d'individus varient selon les lieux, selon les espaces (Cf. 2.1.2.) et le développement par « en bas » dans cette perspective nous interpelle à prendre en compte le facteur local.

En France, ce courant de pensée a conduit au développement d'un certain nombre de travaux comme ceux de PIATIER (1979) qui portent sur la délimitation de l'espace de vie ou ceux de PERRIN (1979) qui mettent l'accent sur l'animation communale. Aux Etats-Unis, on parlera « d'agropolitan development » à la base de la solidarité communaliste et d'autosatisfaction des besoins.

La présente approche poserait implicitement le problème de l'emploi, par le truchement du système économique de chaque espace ou chaque unité spatiale définie. Certes,

on peut saisir ce problème et plus précisément celui posé par l'application des politiques de l'emploi à l'échelle macro-économique. Et dans ce cas, de manière similaire à l'approche précédente, il faudra intégrer les stratégies locales pour atténuer l'effet de leur inefficacité.

2.2.1.3 Approche locale des politiques de l'emploi

En France, la politique de l'emploi a très souvent été considérée de manière centralisée, c'est-à-dire sous un aspect macro-économique. Une telle perspective occulte les contraintes ou les aspirations locales qui ne peuvent que diminuer l'efficacité des mécanismes globaux. Il n'est pas donc étonnant d'écouter le ministre J. Le GARREC lors d'un discours soutenir que « la politique de l'emploi est à la fois macro-économique et micro-économique, globale et démultipliée[...]. La bataille pour l'emploi ne peut se gagner d'en haut. La décentralisation permet de mieux mobiliser les acteurs. Les initiatives locales peuvent et doivent se développer» [36]. On comprend ainsi mieux les stratégies locales de l'emploi, lesquelles s'articulent autour des problèmes que pose l'organisation des politiques de l'emploi dans une perspective macro-économique, et la nécessité de son approche au niveau des territoires fins où apparaissent de nouvelles initiatives depuis la crise. Cette nécessité est d'autant plus urgente que c'est au niveau de ces territoires fins que les réalités socio-économiques sont subies : «c'est à ce niveau que les tensions sociales engendrées par les déséquilibres sont perçues»[37]. Il est alors normal que les élus locaux multiplient les interventions locales en matière d'emploi. Un sondage effectué après les élections municipales de mars 1977 indique que 63 % des maires estiment que l'emploi est leur premier problème et que 52 % souhaitent devenir les véritables responsables de la situation de l'emploi dans leur commune[38]. D'ailleurs la politique d'aménagement des municipalités au cours des années 1960-70 s'est peu transformée en une politique locale d'emploi. C'est donc dans ce contexte que le local émerge, c'est le lieu où s'expriment les solidarités, où se font les pressions sur les élus et que se développent les notions comme les bassins d'emplois ou le marché local, ces notions répondant à la nécessité d'un quadrillage fin du territoire et d'une intervention en groupes sociaux, élus et administrations [39], dans la mesure où il interroge sur le niveau où sont prises les décisions. On comprend alors que la notion de marché local est au cœur des problèmes, et qu'il est indispensable de rendre compte des rapports entre le marché du travail et l'espace, c'est-à-dire intégrer celui-ci dans le fonctionnement de celui-là.

Tous ces problèmes montrent que la réalité du marché du travail, ses caractéristiques et son fonctionnement sont perçus à une échelle géographiquement limitée. LEIDRUT approuve d'ailleurs que si « le marché du travail est réellement le complexe de facteurs économiques et sociaux, à travers lesquels des employeurs recrutent des travailleurs et des travailleurs recherchent un emploi, on ne peut oublier que ces relations se situent dans les limites géographiques peu étendues, constituent un milieu restreint » [40].

L'intérêt d'intégrer l'espace dans le fonctionnement du marché du travail est alors manifeste dans les passages que nous venons d'aborder. Par ailleurs, c'est au niveau des espaces réduits que certains phénomènes, approchés globalement, peuvent permettre d'éclairer les choix nationaux. C'est le cas notamment des problèmes de l'emploi : « il n'y a pas de marché national de l'emploi : celui-ci n'est que le résultat du fonctionnement d'un ensemble de marchés locaux » [41]. Partant de ce principe, nous allons analyser comment les modèles macro-économiques ont pris en compte les phénomènes localisés.

2.2.1.4 L'espace à travers les modèles macro-économiques

L'espace dans les modèles macro-économiques, intervient souvent comme une nécessité d'apprécier l'incidence des facteurs spatiaux sur le développement national et ceci, dans la perspective de la planification. En effet, les décisions de politique économique ou sociale s'appliquant à l'ensemble du territoire national ne sont pas neutres du point de vue spatial. L'espace sera mis en relief à partir des modèles dits mono-régionaux (ils correspondent à une seule région) ou interrégionaux (ils impliquent une multitude de régions et ont l'avantage de mesurer les relations possibles entre les régions). Dans un souci de clarté des analyses que nous poursuivrons ultérieurement, nous allons nous intéresser à ce second type de modèles. En s'inspirant d'une distinction empruntée à BAUVOIR et JAUMOTTE [42] on peut concevoir deux types de modèles interrégionaux : les modèles nationaux à répartition régionale et les modèles régionaux à cohérence nationale.

2.2.1.4.1 Modèles nationaux à répartition régionale.

Il a été déjà indiqué que certains modèles macro-économiques souffraient de leur inadaptation aux phénomènes localisés. C'est dans l'optique de surmonter de tels problèmes qu'on a été conçus les modèles nationaux à répartition régionale. Il s'agit de régionaliser certaines variables du modèle national, de façon à prendre en compte les spécificités régionales voire locales. C'est ainsi qu'au niveau du modèle global, on peut remplacer l'emploi total par sa ventilation suivant les régions. NEGRE [43] distingue trois techniques de régionalisation :

- *la technique de répartition exogène* qui consiste à désagréger la variable nationale sur la base d'un système de pondération dont la somme est égale à l'unité ;

- *la technique de répartition "à rétroaction limitée"* par laquelle la variable nationale est décomposée grâce à l'utilisation des équations de comportement pour les ratios de distribution ;

- *la technique de régionalisation complète* qui consiste à remplacer l'équation nationale par les équations régionales.

Le premier constat à faire ici est que l'espace intervient de façon exogène au modèle. Il n'émane pas des pratiques des agents économiques et on revient presque au concept d'espace « contenant un contenu ». Nous allons néanmoins progresser en analysant comment le marché du travail intervient dans un tel modèle. Nous l'illustrerons sur l'exemple du modèle belge RENA [44] constitué de deux blocs : le bloc national (il comprend le revenu, la demande, les prix les équations monétaires) et le bloc régional centré sur le marché du travail et la fonction d'investissement.

Dans ce modèle, la régionalisation se fait par le biais du P.N.B. dont la ventilation régionale permet de déterminer la capacité de production de chaque région, laquelle induit une demande de main-d'œuvre qui, confrontée à l'offre de travail permet de mesurer le chômage régional. A partir de ce niveau de chômage, et le niveau des prix déterminés dans le bloc national, on fixe le salaire de chaque région. Certes, il y a un dialogue nation-région ou

pour simplifier, il y a prise en compte du facteur spatial dans le modèle. Cependant l'espace intervient comme un cadre dans lequel on confronte l'offre globale de travail à une demande globale de travail. On retombe donc au schéma du modèle traditionnel du marché du travail où les caractéristiques du marché n'interviennent pas. L'espace s'impose ici comme un facteur de séparation. Le phénomène d'interrelation spatiale est moins perçu dans les modèles nationaux à répartition régionale, ce qui n'est pas le cas pour les modèles régionaux à cohérence nationale.

2.2.1.4.2 Modèles régionaux à cohérence nationale.

Par rapport aux précédents modèles, les comportements régionaux, de même que les interdépendances spatiales sont largement pris en compte dans les modèles régionaux à cohérence nationale. Ainsi, la région à travers l'espace n'est pas considérée comme une entité indépendante. On constate donc à ce niveau qu'il y a un net progrès par rapport à la perspective des modèles nationaux à répartition régionale. Mais les interdépendances régionales matérialisent-elles un certain fonctionnement des marchés du travail ? La question mérite quelques éléments de réponse. Au préalable, il faut noter que les interdépendances spatiales interviennent lorsque l'on suppose par exemple que le niveau de production d'une région dépend pour certains secteurs d'activité de la production, il est clair que toute modification de la production dans l'une des régions aura des répercussions spatiales en termes d'emploi. Toutefois, ces répercussions ont lieu dans un cadre bien défini et ne semblent pas déborder de ce cadre préétabli (la région est dans ce cas une entité administrative).

On peut déceler derrière cette approche, une propension à prendre en compte des éléments structurels – le fait que la production et par conséquent l'emploi soient abordés par secteur – mais qui ne peut être valablement saisi par le modèle, du fait de son caractère a-structurel. Ainsi, bien qu'il y ait des liaisons entre les régions, l'indicateur d'ensemble qu'on privilégiera sera l'emploi global au niveau d'une région.

Nous allons au moyen du modèle REGINA[45] qui est l'exemple type des modèles régionaux à cohérence nationale, voir comment en intégrant l'espace, on aborde le fonctionnement du marché du travail. Le modèle va essayer d'articuler les deux blocs qui le constituent : le bloc régional et le bloc national.

Le bloc régional est construit sur la base de la différenciation spatiale des entreprises en matière de localisation et les contraintes géographiques qui pèsent sur elles. Pour simplifier, il existe trois types d'activités selon les facteurs qui président à leur localisation ; elles sont réunies dans chaque bloc régional. Celui-ci est appréhendé par un tableau d'échange interindustriel régional (T.E.I.R.) qui décrit les relations intra et interrégionales.

Il existe deux phases dans l'articulation du modèle :

1^{ère} phase : on détermine les perspectives nationales à partir d'hypothèses régionales. Prenons le cas de l'évolution des salaires au niveau national. Elle dépend de la dispersion du chômage sur les divers marchés régionaux de l'emploi (cette hypothèse fait implicitement appel à la courbe de Philips) ; le taux de décroissance moyen des salaires est d'autant plus élevé au plan national que le chômage est fortement dispersé entre les régions[46].

2^{ème} phase : on régionalise les perspectives nationales. Connaissant les valeurs de la production par branche et par région, lesquelles sont déterminées par confrontation du T.E.I national et des T.E.I.R., on peut, sous l'hypothèse d'une connaissance de la productivité apparente du travail dans chaque région, prévoir la demande régionale du travail qui sera confrontée à l'offre régionale du travail. Le modèle suppose que tout déséquilibre sur le marché du travail sera résorbé par les migrations interrégionales. On retrouve le paradigme néoclassique du marché du travail. L'espace est bien sûr pris en compte dans le modèle, mais il ne résulte pas non plus de l'action des agents économiques. C'est encore un lieu où l'on confronte une offre et une demande globale de travail.

Au-delà de ces premières observations liées au marché du travail, on peut dire que les deux types de modèles que nous venons d'étudier sont « inaptes à l'analyse du système des rapports sociaux sous-jacents aux mécanismes économiques pris en compte ». Autrement dit, ils ne peuvent mettre en lumière les changements structurels qui interviennent sur le marché du travail. Peut-être ceci est-il dû au fait que ces modèles s'appuient sur la comptabilité nationale, qui définit des agents par rapport à des fonctions économiques et non des pratiques

sociales ? Ainsi les entreprises sont définies comme le lieu de la formation de la valeur ajoutée et de la distribution des revenus primaires et non comme le lieu des rapports conflictuels entre employeurs et salariés. Dans une perspective plus analytique nous verrons comment l'espace est intégré dans le modèle concurrentiel. Tel est l'objet du second paragraphe.

2.2.2 L'espace dans le modèle traditionnel du marché du travail.

Le modèle traditionnel du marché du travail est par essence a-spatial. Cependant, il vient d'être observé dans les précédentes analyses que l'espace n'est pas neutre et qu'il jouerait un rôle fondamental dans la compréhension des mécanismes de fonctionnement du marché travail. En effet, la projection du modèle dans l'espace montre qu'il y a inadéquation entre le vécu du marché du travail et le schéma proposé par la théorie. Il y a donc remise en cause du modèle : « le propre du facteur spatial [...] est de remettre en cause la signification des catégories traditionnelles et les postulats fondamentaux de l'analyse » [47]. Le problème qui se pose alors est de savoir comment intégrer l'espace dans le modèle de manière à adapter la théorie aux faits. Selon LESTER « toute théorie qui ne réussit pas à se concilier les faits doit être écartée comme mal fondée ou modifiée de façon à fournir une explication adéquate ». L'intégration de l'espace dans le modèle traditionnel peut donc être perçue comme une tentative de modification du modèle traditionnel pour se rapprocher de la réalité. Mais cette perspective est-elle assez pertinente pour l'analyse ? En d'autres termes, s'agit-il d'une approche permettant d'aborder l'espace comme un cadre à l'intérieur duquel on applique les outils du modèle traditionnel (espace géonimique) ou bien, est-il défini par les structures qui le façonnent ? Nous l'apprécierons à travers le développement qui suit et où nous intégrerons l'espace dans le modèle concurrentiel sur la base soit d'un critère, soit d'une modification de quelques hypothèses de base. Nous tenterons au mieux d'une modification de quelques hypothèses de base. Nous tenterons au mieux d'apporter des éléments critiques aux problématiques induites par de telles approches.

2.2.2.1 Spécialisation du marché concurrentiel sur la base des critères.

Tout en faisant abstraction à la dimension géographique du marché, a déjà observé quelques limites du modèle traditionnel (Cf. 1.1.3), lequel était incapable d'expliquer des différences apparues à partir d'une époque sur le marché du travail. Sur ce point, on ne peut pas nier les apports néoclassiques pour surmonter ces limites afin de concilier le modèle et les faits, notamment par le biais des « théories locales » (théorie du capital humain, théorie de la recherche d'emploi ou « job-search », théorie de la discrimination). Toutefois, ces théories expliquent les différences à l'intérieur du marché par les facteurs qui sont totalement exogènes à son fonctionnement.

L'approche spatiale du modèle concurrentiel repose presque sur le même principe. Ainsi peut-on sur la base des critères définir un marché auquel associe l'espace. Les critères tels que la qualification, l'âge, le sexe, la profession etc. ont des rapports certains à l'espace. Les études empiriques montrent que le marché d'un cadre a une dimension nationale alors que celui d'un salarié peu qualifié se situe à une échelle. Cette dichotomie est relativement similaire à la dualisation spatiale du marché du travail étudiée par PLANQUE et qui oppose le marché du travail des salariés qualifiés dans les centres urbains à celui des salariés peu qualifiés dans les régions périphériques.

En se basant également sur les critères, BENHAYOUN [48] montre que la dimension urbaine intervient de manière très significative sur le marché des jeunes. Cette thèse n'est pas très loin de celle de H. JAYET [49] qui trace une certaine similitude entre le marché de travail des femmes et celui des jeunes. On peut dire que sur la base des critères, il est possible de définir le marché des jeunes, le marché des cadres, le marché des femmes, etc.. Il s'agit évidemment d'une approche unidimensionnelle, où l'espace sert comme un cadre à l'intérieur duquel seront appliquées les hypothèses du modèle concurrentiel, dans la mesure où les catégories définies sont supposées homogènes. L'hypothèse d'homogénéité apparaît d'ailleurs invraisemblable. Parler par exemple de marché de marché des jeunes a-t-il un sens lorsque l'on sait, ainsi que l'indique P. SARTIN qu'«il existe au sein de la jeunesse de profondes divergences dans les comportements qui varient avec l'âge, la prise de conscience des réalités de la vie professionnelle, le degré de maturité psychique, le sexe, le milieu social, voire la région » ? [50]. La jeunesse n'est donc pas un groupe homogène parce qu'il est difficile de l'isoler des autres groupes d'âges, et parce qu'elle est traversée de multiples

différenciations. Prendre le critère d'âge pour délimiter le marché dans l'espace n'est pertinent pour l'analyse. La spatialisation du marché sur la base des critères est manifeste dans les travaux de planification où, « il s'agit alors de fixer les conditions de l'équilibre, les structures de l'emploi dans l'espace dont l'origine est à rechercher dans la logique même du plan. Il est alors purement exogène par rapport au fonctionnement du marché du travail ». L'approche n'apporte donc aucune autonomie au marché ; l'espace est mal limité et apparaît comme le résultat d'un certain fonctionnement qu'il tente de mieux représenter.

L'espace tel qu'il est perçu ici, s'oppose à la notion qui a été définie plus haut et qui repose sur les pratiques voire les stratégies des agents, lesquelles ont pour faculté de structurer géographiquement un marché ; celui-ci est alors endogène à son fonctionnement. Cette deuxième perspective est donc celle qui prend en compte les différentes structures qui caractérisent un marché du travail. Avant d'y revenir ultérieurement, nous allons voir maintenant comment la modification des hypothèses du modèle concurrentiel peut nous conduire plutôt à intégrer l'espace dans le modèle.

2.2.2.2 Spatialisation du marché concurrentiel sur la base de la modification des hypothèses du modèle.

Lorsque les hypothèses du modèle concurrentiel sont posées, elles supposent un espace neutre qui ne fasse pas obstacle à leur vérification. Cependant, un examen assez détaillé montre qu'elles traduisent un certain rapport à l'espace. Ce sera donc l'objet des analyses qui suivent.

2.2.2.2.1 Mise en évidence du facteur spatial dans l'hypothèse de la mobilité du travail

Dans la perspective du modèle concurrentiel, la mobilité du travail rencontre très rapidement des limites du fait de l'immobilité relative du facteur travail par rapport au facteur capital dans l'espace. Celui-ci sera introduit dans le modèle par le biais d'un coût de déplacement, lié à la distance géographique : la distance lieu de travail – lieu de résidence

implique un coût de déplacement dont la configuration va modeler l'espace du marché, conçu comme la zone dans laquelle la mobilité est possible. Les zones de migrations alternantes s'inscrivent dans une telle perspective et MALLET [51] montre que 80 % de changement d'emploi ne s'accompagnent pas d'un changement de résidence dans de telles zones. On peut alors expliquer les différences de salaire dans un tel schéma par l'existence des coûts de déplacement qui les compensent.

Il est patent que le coût de mobilité lieu de travail – lieu de résidence permet de faire apparaître l'espace dans le modèle. Mais au-delà, il existe des limites :

- le coût ne peut être lié uniquement à la distance géographique. En effet, on a déjà vu que l'individu n'était pas insensible à l'endroit où il travaille, qui présente des avantages ou des inconvénients dépendant surtout des structures de sa localisation. Le climat, les relations, les équipements et les loisirs, etc. . sont des facteurs qui peuvent modifier dans un sens ou dans l'autre les coûts de déplacement, c'est-à-dire rendre le déplacement envisageable pour des différences de salaires plus ou moins importantes pour une même distance parcourue. Ainsi, outre le coût de déplacement, le modèle devrait être à même de prendre en compte le coût de changement d'environnement ;
- la structuration de l'espace résulte des seuils dans les coûts. Peut-on envisager que tous les individus ont un même seuil ? Répondre à la question revient, entre autres à recenser tous les facteurs qui font qu'un individu étende ou pas l'aire de sa recherche d'emploi. Ces facteurs sont loin d'être identiques et une analyse structurelle peut permettre d'y voir plus clair. De plus, structurer l'espace par les seuils pose le problème de stabilité des limites du marché puisque toute fluctuation des coûts aura des effets sur ces limites.

Si l'on suppose homogène une catégorie d'individus, peut-on également dire qu'ils ont un comportement migratoire identique ? Cette hypothèse ne peut être justifiée dans la mesure où les mobilités sont différemment pondérées selon que l'on est célibataire ou père de famille, jeune ou vieux etc. . . et l'on voit là encore qu'une analyse structurelle devient nécessaire.

On peut alors conclure que la dimension spatiale du marché résulte ici de la logique de la localisation spécifique au facteur travail ; l'espace est pris en compte dans le modèle surtout

comme facteur de séparation des différents lieux où le travail est localisé. En sera-t-il le cas lorsque l'on modifie l'hypothèse relative aux conditions de diffusion de l'information ?

2.2.2.2.2. L'espace et les conditions de diffusion de l'information.

Dans le modèle concurrentiel, l'information est supposée d'une facile accessibilité à tous les agents sur le marché. L'espace est alors considéré comme homogène en tant que support de l'information. Cette vision est assez simpliste et l'information n'est pas diffusée de façon homogène en tout point du marché. Les travaux de SUGDEN [52] montrent que la taille du marché influence le degré d'accessibilité à l'information et également le coût de sa recherche. Plus le marché est grand (spatialement), plus la diffusion de l'information est lente et plus les coûts de recherche des informations sont également importants.

L'approche est celle qui associe le coût de l'information à la distance parcourue. En effet, « le rôle d'un marché étant de mettre en contact les acheteurs et les vendeurs et de permettre la diffusion de l'information aux intéressés, chaque marché aura nécessairement une étendue plus ou moins limitée compte tenu des possibilités de transmission de l'information » [53]. La collecte de l'information par le coût qu'elle représente permet d'introduire la distance. Cette vision est différemment partagée par la théorie de la recherche de l'emploi qui soutient que le coût de l'information est plus lié au temps nécessaire à son acquisition qu'à la distance. Même si les moyens traditionnels d'information sur le marché du travail (relations personnelles, petites annonces, services administratif (ANPE) se situent dans des espaces géographiques limités, il n'en demeure pas moins qu'il faut quelque fois « plus de temps pour s'introduire dans certains réseaux de relations que pour téléphoner à une ville éloignée ». La distance géographique n'apparaît plus à ce stade et c'est la plus ou moins grande facilité d'acquisition de l'information qui devient prédominante.

Associer des coûts à l'information revient à fixer des limites dans l'espace, ainsi qu'on l'a vu avec les zones de migrations alternantes. Il s'agit certes d'une amélioration du modèle concurrentiel dans le but d'intégrer l'espace, mais cette démarche fait table rase sur les réseaux de mise en relation, qui eux ne dépendent pas de la distance : « l'information qui dans

le passé, circulait moins vite et moins loin permettait de limiter les zones ; lorsqu'une information se diffusait de A vers C, les canaux qu'elle empruntait amenaient presque toujours une diffusion continue entre A et C aujourd'hui certaines nouvelles vont plus facilement de Paris à New York que de Paris en Bretagne. Les moyens techniques eux-mêmes (avions...) favorisent une diffusion discontinue et donc sélective. La notion de délimitation géographique perd alors de son intérêt. On retrouve l'idée de réseau » [54].

2.2.2.2.3 Prise en compte de l'espace à travers l'hypothèse d'atomicité des agents

Dans le modèle traditionnel du marché du travail, l'espace devient neutre dès lors que l'on suppose que tout agent sur le marché a le même pouvoir que les autres. Projetée dans l'espace, cette hypothèse n'est pas assez tenable (les différences spatiales interviennent) et on a essayé de la modifier afin d'intégrer la dimension spatiale dans le modèle. Cette transformation de l'hypothèse est également basée sur les coûts de déplacement comme dans les analyses que nous avons faites dans les deux sous-paragraphes précédents. Nous nous situons ici dans la perspective des marchés interdépendants, ce qui nous démarque un peu de celle abordée en (2.2.2.1.) où les marchés étaient limités par les coûts de déplacement et apparaissent comme indépendants les uns des autres.

On peut ainsi concevoir, en ce qui concerne les individus, que certains ont des positions géographiques leur permettant d'intervenir sur plusieurs marchés pour offrir leur travail. Du fait de cette situation privilégiée, ils peuvent avoir à négocier des contrats de travail en plusieurs lieux, ce qui ne veut pas dire qu'ils travaillent en plusieurs endroits à la fois. Ils ont donc un champ d'opportunités plus vaste ; leurs choix pour l'emploi sont nombreux. A l'inverse, les individus moins bien situés géographiquement devraient pour bénéficier des mêmes opportunités que les précédents, supporter des coûts de déplacement plus importants pour être sur autant de lieux de négociation. On voit que l'introduction de l'espace facteur de coût modifie l'hypothèse d'identité des pouvoirs entre les individus, ce qui est bien entendu une façon de prendre en compte l'espace dans le modèle. Cependant, le critère coût de déplacement ne saurait être pertinent pour cerner l'aire de recherche d'emploi des individus et nous l'avons déjà mentionné plus haut.

Si l'on se place du côté des entreprises, on constate que leur rôle est très limité dans le modèle. Celui-ci est construit sur la base de petites entreprises, lesquelles ne peuvent se prévaloir d'utiliser l'espace. Elles ne peuvent donc se déplacer facilement, car les coûts de déplacement seraient assez importants. Leur pouvoir limité sur l'espace ne se manifestant qu'au niveau de la localisation, elles ont alors intérêt à « toujours se situer là où les salaires sont les plus bas, et (d') éviter (contrairement aux individus) les zones de recouvrement entre marchés qui peuvent les (la) contraindre à des augmentations de salaires ». L'introduction de l'espace comme coût n'induit pas de comportements particuliers de la part des entreprises déjà implantées ; la perspective abordée ici reste basée sur les comportements des individus uniquement. Le rôle structurant de la demande de travail n'est pas mis en relief, ce qui constitue là une limite d'approche évidente.

On peut conclure ici en observant que le facteur spatial est un élément fondamental dans l'étude du fonctionnement du travail. Certes, les analyses que nous venons de faire montrent la nécessité de le prendre en compte, mais des limites subsistent dans la mesure où l'espace intervient de manière exogène. Il s'agit d'un recours à l'espace géonimique, perçu comme contenant un contenu. Au-delà, nous allons voir si l'exemple de quelques marchés concrets s'inscrit dans cette perspective.

2.3 ETUDE DE QUELQUE CAS DE MARCHE LOCALISE

L'analyse que nous entreprendrons dans la présente section nous conduira à examiner d'abord les exemples de marchés locaux (2.3.1.) et ensuite d'évoquer les problèmes posés par la détermination des limites géographiques des marchés.

2.3.1 Exemples de marchés locaux.

Dans ce paragraphe, nous présenterons quelques exemples de marché locaux à savoir : les bassins d'emploi, les zones d'emploi, l'exemple des bassins de main-d'œuvre d'Annecy et de Compiègne.

2.3.1.1 Bassins d'emplois.

Le concept de bassin d'emploi, né de la pratique de l'aménagement du territoire, est apparu autour des années soixante-dix dans de nombreuses études sociologiques [55]. Il est de nos jours très couramment utilisé par les responsables administratifs, les élus nationaux voire locaux. Avec la pérennité de la crise amorcée depuis 1974, son utilisation est de plus en plus manifeste lors des tentatives de régulation locale par des acteurs locaux. C'est donc à ce niveau que sont profondément ressenties les tensions induites par les déséquilibres sur le marché du travail.

Le concept de bassin d'emploi peut ainsi s'appuyer sur le désir des populations de travailler et de vivre au « pays », ce que tend à confirmer le courant de recherche basé sur le développement « par en bas » (Cf. 2.2.1.2.), qui prend prioritairement en compte les besoins collectifs des communautés localisées. Ces premiers éléments nous conduisent à l'ébauche d'une définition de ce concept : c'est la zone où les gens trouvent « pour la plupart » leur travail et où les entreprises trouvent la « plupart » de leurs employés ; c'est donc la zone la plus adéquate pour recueillir une information pertinente sur le marché du travail et pour mener une politique locale de l'emploi » [56]. Rien de plus étonnant que les études lui faisant référence depuis la crise visent à son approfondissement théorique et pratique afin d'aboutir à des actions concrètes.

Au-delà de cette définition, il apparaît des rapports des agents économiques à l'espace : une interaction existe entre les entreprises et le bassin d'emploi, celui-ci étant porteur d'une offre de travail que les entreprises recherchent et « trouvent ». On est alors à même de se poser la question à savoir si les bassins d'emploi peuvent former un espace économique. Pour GAMBIER, les bassins d'emploi sont des entités économiques capables d'adaptation, mais « cette faculté semble hypothétique dans la mesure où un certain nombre de bassins d'emploi n'ont pas une structure économique diversifiée pour amortir les aléas de la conjoncture » [57]. Les bassins d'emploi sont donc liés à l'activité du marché local du travail et suivant les fluctuations conjoncturelles (ouverture ou fermeture d'une grande entreprise) ou l'amélioration des infrastructures ; leur dimension peut se modifier.

2.3.1.2 Les zones d'emploi.

L'utilisation de la notion de zone d'emploi a été systématisée et généralisée à partir de 1982 quand le ministère des affaires sociales et de la solidarité nationale a voulu chercher un découpage pertinent pour l'étude de l'emploi. Il faut noter que le découpage administratif existant auparavant ne permettait plus aux différents auteurs socio-économiques d'avoir une vision fine et pertinente des problèmes locaux de l'emploi[58] ; ils n'étaient pas toujours adéquats pour l'analyse des relations économiques. Il a été ainsi défini 365 zones d'emploi sur tout le territoire, et elles sont basées sur le critère d'activités économiques. Ce zonage reste le fruit d'une procédure décentralisée et sa détermination se fait suivant un compromis satisfaisant entre les exigences variées et parfois contradictoires : respecter les découpages administratifs régionaux existants (elles respectent partout les frontières régionales), avoir une dimension suffisante pour permettre la collecte statistique, avoir une signification économique (chaque zone doit servir de point d'appui aux études économiques et doit traduire la répartition géographique des activités économiques et l'intensité plus ou moins forte des liens qu'elles tissent entre les points du territoire) et respecter par conséquent les navettes domicile-travail. On voit bien nettement ce qui pourrait différencier ou rapprocher cette notion de celle des bassins d'emploi.

La zone d'emploi se définit ainsi comme un « espace géographique à l'intérieur duquel les habitants trouvent normalement un emploi, et dans lequel les établissements trouvent la main-d'œuvre nécessaire en qualité et en quantité pour occuper les emplois qu'ils procurent : l'aire de recrutement y coïncide avec l'aire d'emploi » [59].

Ce découpage reste tout de même celui d'un espace limité géographiquement et laisse subsister quelques problèmes. En effet, il est basé sur l'activité économique et ne peut pas être figé, du fait des mutations voire des restructurations industrielles qui peuvent façonner et rendre ainsi obsolètes les critères qui ont prévalu lors du découpage initial. D'autre part, ces zones doivent respecter les navettes domicile-travail. Or celles-ci, comme la plupart des mouvements économiques dans l'espace traduisent d'abord « l'adaptation des agents économiques à des contraintes d'une période » et les zones d'emploi peuvent ne refléter qu'imparfaitement dans l'avenir les phénomènes économiques qui ont fondé leur détermination. Ceci reste également vrai pour ce qui est des bassins d'emploi. Sur un plan

analytique, associer ces zones à l'espace du marché du travail revient à occulter les relations mutuelles qu'elles peuvent entretenir. On revient implicitement à la logique d'un cadre fermé à l'intérieur duquel on applique les outils du modèle concurrentiel.

2.3.1.3 Marchés locaux concrets

Certaines recherches du centre d'études de l'emploi (C.E.E.) ont conduit à l'étude de quelques marchés, notamment dans les régions d'Annecy et de Compiègne. L'objectif pour les auteurs de ces études, de descendre à ce niveau géographique réduit, consiste à « cerner » :

- toutes les opérations effectuées sur le marché du travail par les employeurs, d'une part, les salariés, d'autre part ;
- les flux résultant de ces opérations » [60].

Ainsi, sur une période donnée, il a été possible de décrire l'ensemble des mouvements caractérisant l'activité d'un marché du travail dans ces régions. Celles-ci sont délimitées sur la base des migrations alternantes : elles ont peu de migrants alternants avec les zones géographiques limitrophes.

Les deux régions constituent des marchés du travail au sens de la définition empruntée à ROBINSON : « un marché du travail est la région qui contient les membres effectifs ou potentiels de la force du travail qu'une entreprise pourrait amener sous certaines conditions, à prendre un emploi, et les autres entreprises avec lesquelles elle se trouve en concurrence pour la main-d'œuvre » [61]. On note ici qu'il s'agit d'une délimitation renvoyant à l'espace limité par les frontières continues, contenant tout ce qui est à l'intérieur des frontières et rien de ce qui est à l'extérieur » [62]. Elle nous ramène aux critiques que nous avons déjà évoquées auparavant. Il serait juste d'indiquer l'importance accordée à l'entreprise dans cette définition, ce qui est une façon de mettre l'accent sur son rôle structurant qui nous conduit aux thèses «segmentationnistes ».

Le marché du travail déterminé sur le bassin d'Annecy est décrit par une comptabilisation de flux de main-d'œuvre. On a pu y mesurer la mobilité des travailleurs et

préciser quelques déterminants dans le fonctionnement du marché du travail. Il apparaît une segmentation autour de deux sphères : la sphère des mobiles et la sphère des stables. Cette segmentation est un processus désignant l'hétérogénéité des modes de gestion de la main-d'œuvre et des comportements individuels. Ces deux sphères semblent néanmoins assez globales et méritent d'être affinées suivant les critères tels que l'âge, le sexe, la qualification, etc. .

C'est sur la base des travaux sur le bassin d'Annecy qu'a été élaboré un cadre du marché du travail, auquel se rattachent les études menées sur le bassin d'emploi de Compiègne. Ici, il apparaît que la mobilité du personnel ou sa stabilité est pour une large part liée à la structures des entreprises et non à leur leurs attractivités (salaires, conditions du travail). Par ailleurs, les fréquences de mobilité sont nettement liées à l'âge et au mode de participation à la vie active, ce qui rend plus pertinente l'information qu'on tire de la connaissance des sphères sur lesquelles est centrée la segmentation. Ainsi, la sphère des instables serait liée à ces deux variables.

Les différents marchés locaux qui viennent d'être proposés s'appuient sur des espaces limités à l'intérieur desquels fonctionne de façon autonome un marché. Dans cette optique, sur quelles bases détermine-t-on les limites de ces marchés ?

2.3.2 Détermination des limites géographiques du marché local.

Le concept de marché de travail tel que nous l'avons décrit jusque maintenant est celui d'un espace géographique limité. Même si ce point nous semble très restrictif, il semble présenter les démarches qui conduisent aux délimitations géographiques du marché local. GAMBIER distingue deux types de démarches [63] : les démarches empiriques et les démarches statistiques.

2.3.2.1 Démarches empiriques.

La détermination des limites du marché du travail local sur la base des démarches empiriques reposera sur les découpages administratifs, la pratique de certains acteurs ou encore sur la notion de pays.

Le découpage administratif s'inscrit dans le cadre de nombreuses interventions de l'administration, notamment en matière d'emploi. Très généralement, ce sont les migrations alternantes qui fixent les contours des espaces recherchés. On peut citer à titre d'exemple : les ZPIU (zone de peuplement industriel et urbain), les zones d'aides à l'aménagement du territoire, et les zones d'emploi de l'ANPE.

Les ZPIU sont constitués d'un regroupement de communes caractérisées par une certaine homogénéité de peuplement, d'importants échanges de population entre elles et une activité industrielle réelle. Ce sont ces caractéristiques qui ont amené certains auteurs à les assimiler à des bassins d'emploi ;

Les zones d'aides à l'aménagement du territoire forment un autre type de zonage défini par la DATAR sur la base du taux de chômage. Il s'agit ici d'un espace à l'intérieur duquel l'administration cherche à établir un certain équilibre du marché du travail socialement plus acceptable, et telle semble l'idée sous-jacente à la construction d'un tel espace, qu'on associe au marché local. De plus, ce zonage a une fonction surtout politique ; celle-ci se traduira par des avantages fiscaux ou des aides financières aux entreprises s'y trouvant, de façon à permettre la réalisation de l'objectif recherché ;

Le découpage de l'ANPE en zones locales de l'emploi où chacune correspond à une antenne locale de l'agence et ayant pour but d'améliorer les résultats du placement et la gestion de l'agence. Ce zonage permet de produire à un niveau fin l'information sur le chômage et a contribué à l'émergence de la notion de bassin d'emploi.

On peut également définir le marché local du travail à partir de la *notion de pays* : « le pays est toujours une aire spatiale traditionnelle, de dimensions limitées, plus petites que les départements..., reconnue par les habitants comme entité spatiale originale différente des pays voisins et commandant un sentiment d'appartenance » [64] ; c'est « la zone géographique dans laquelle la quasi-totalité des hommes sont à la fois habitants, producteurs et consommateurs » [65]. On se situe dans le paradigme de développement d'écosystèmes de petite taille (agropolitan development) mis en relief par FRIEDMAN et DOUGLAS [66]. La notion de pays relève plus du vécu, des traditions, de la dimension historique et culturelle et

ne s'appuie pas sur les forces économiques. Le pays et le lieu de recherche d'une unité de gestion sociale et c'est dans ce sens qu'on le définit comme un marché local.

Le marché local peut naître également de la pratique de certains acteurs comme des organisations professionnelles qui souvent localement jouent un rôle fondamental en matière d'emploi. C'est le cas par exemple des fédérations syndicales, qui ont des zones d'intervention où leur action s'étend aussi bien sur les recrutements que sur la formation etc..

Il est possible de multiplier les exemples de marchés locaux relevant donc de l'approche administrative ou pour simplifier empirique. Dans tous les cas, on constate que les limites s'imposent au cadre spatial retenu et ne découlent pas du fonctionnement même du cadre.

2.3.2.2 Démarches statistiques.

L'essor de l'informatique et les possibilités de traitement qu'elle entraîne ont permis de mettre au point des méthodes statistiques servant à découper les marchés locaux. L'I.N.S.E.E. a ainsi pu développer des méthodes s'apparentant au procédé de classification hiérarchique ascendante, lesquelles sont basées sur des flux de main-d'œuvre entre points du territoire. On peut citer la méthode baptisée ZONAGE entreprise par la Direction Régionale de l'I.N.S.E.E de Rennes[67] ou encore celle dénommée MIRABELLE (Méthode Informatisée de Recherche et d'analyse des Bassins par l'Etude des Liaisons Logement-Emploi) et développée à la Direction Régionale de l'I.N.S.E.E. de Nancy [68]. Par ces méthodes, on déterminera par exemple le bassin d'emploi comme l'espace à l'intérieur duquel s'effectue la plus grande partie des déplacements domicile-travail.

Avec MIRABELLE, on recherche non seulement à obtenir un découpage de l'espace mais une succession de découpages emboîtés les uns dans les autres. Pour schématiser, on suppose deux zones quelconques. Elles seront d'autant plus fortement liées que le nombre de migrants alternants qui leur est associé est élevé. Dans le cas où les deux zones sont alors liées, on les agrège et on recalcule le lien entre l'agglomération ainsi constituée et chacune des autres zones ; on cherche de nouveau les deux ensembles les plus liés et on recommence

jusqu'à couvrir totalement une région voire le territoire. Cette procédure permet d'apprécier la dynamique qui se fait jour à l'intérieur des bassins et entre eux, d'une part, par l'ordre dans lequel s'agrègent les zones, d'autre part, par l'intensité des échanges qui les lient. Il faut noter que le passage d'un bassin à l'autre se fait au-delà d'un certain seuil qu'on se fixe (10 % généralement). Autrement dit, toute modification de ce seuil remettra en cause les limites des bassins constitués. Il y aura alors autant de limites que de seuils retenus dans la méthode, ce qui rend la définition des bassins assez subjective. Nous ne revenons pas ici aux critiques déjà faites auparavant sur le critère des migrants alternants. Il sera abordé maintenant une approche statistique du marché local développée par GIARD où l'auteur nous propose deux méthodes pour le découpage spatial.

2.3.2.3 Marché local dans la perspective de GIARD

Dans ses travaux, GIARD présente un modèle de « prévision de l'emploi spatialisé » qui propose quelques instruments pour situer l'emploi dans l'espace. Selon l'auteur, « le marché de l'emploi est défini par l'ensemble des trois caractéristiques suivantes, qui régissent d'ailleurs les unes sur les autres : un cadre temporel, un cadre spatial et un ensemble de mécanismes ».

Le cadre temporel, c'est le moyen terme et l'ensemble des mécanismes, c'est « les interactions et déterminations des variables qui définissent l'équilibre à un instant donné ». Il reste à définir le cadre spatial et nous allons montrer comment l'auteur le détermine. S'appuyant sur le concept d'espace homogène, il utilise deux méthodes statistiques permettant de choisir le meilleur découpage parmi ceux qui existent ; la commune, le canton, la ZEDE (zone d'Etudes Démographiques et d'Emploi) le département. Il s'agit des méthodes, l'une basée sur les migrations entre zones, l'autre sur la comparaison de quelques variables de structure.

2.3.2.3.1 Méthode fondée sur les migrations entre zones.

V. GIARD construit sa méthode sur le critère de minimisation des migrations définitives entre zones. Il préfère ainsi les migrations définitives aux migrations alternantes parce qu'il veut prendre en considération les zones pour lesquelles « il existe une transparence satisfaisante des conditions d'offre et de demande d'emplois ». La définition du cadre spatial revient donc à choisir le meilleur regroupement d'unités spatiales de base.

Le principe est le suivant : on détermine un indicateur de rétention d'une zone quelconque issue d'un découpage. Cet indicateur représente la part des migrants restés dans la dite zone. Par une certaine transformation, on aboutit à un coefficient de rétention global ; on retient le coefficient de rétention global le plus élevé pour un découpage donné et qui traduit le meilleur découpage.

Plus formellement, soit r^* l'indicateur de rétention de la zone T^* . On suppose que T est l'ensemble du territoire découpé en un nombre fixe d'unités spatiales de base ; T^* représente les divers regroupements possibles.

$$r^* = \frac{MIG(T^*, T^*)}{\sum_T MIG(T^*, T)} \quad 0 \leq r^* \leq 1.$$

où $MIG(T^*, T^*)$ est l'effectif de migrants ayant changé d'unité spatiale à l'intérieur de la zone T^* .

$\sum_T MIG(T^*, T)$ est l'effectif de migrants ayant changé d'unité spatiale à partir de toutes les unités spatiales de la zone T^* et à destination de toutes les unités spatiales du territoire.

En pondérant les coefficients de rétention par la population de la zone T^* qu'ils caractérisent, on obtient le coefficient de rétention globale K défini par :

$$K = \sum r^* \frac{Population\ T^*}{\sum Population\ T^*}$$

GIARD a appliqué ce principe à la région Nord, et c'est la ZEDE, correspondant au découpage de la région en 17 zones, qui assure le meilleur découpage. Le coefficient de rétention global est de 0,68.

L'application de cette technique montre une fois de plus qu'il s'agit de la recherche d'un cadre homogène adapté aux mécanismes globaux d'ajustement, c'est-à-dire un cadre où l'on peut identifier les marchés de type néoclassique. Par ailleurs, la similitude est assez manifeste ici avec les modèles macro-économiques que nous avons explorés (RENA ET REGINA). D'autre part, l'indicateur de rétention peut-il permettre de mesurer efficacement les limites entre zones ? Ceci ne semble pas le cas puisqu'il ne fait pas intervenir l'immigration. Ainsi, cet indicateur peut être égal à l'unité dans une zone qui «importe» beaucoup de main-d'œuvre.

Une seconde méthode est proposée par l'auteur pour effectuer des découpages spatiaux, notamment celle qui reste fondée sur les variables structurelles.

2.3.2.3.2 méthode fondée sur les variables structurelles.

Alors que dans la première méthode, GIARD utilise un critère d'ouverture relative des zones les unes sur les autres, dans la deuxième il s'appuie sur des variables de structure qu'il compare. Ces variables au nombre de quatre, sont les suivantes :

- le taux de chômage des 20-65 ans
- le taux d'activité global des 20-65 ans
- le taux d'activité féminin des 20-65 ans
- le coefficient de stabilité rurale :
$$\frac{\text{Actifs agricoles en } t}{\text{Actifs agricoles en } (t - 1)}$$

Le meilleur découpage sera celui qui minimise les disparités spatiales entre ces indicateurs, c'est-à-dire celui pour lequel les différences internes sont nettement plus faibles que les différences avec l'extérieur. Les résultats obtenus par l'auteur montrent que le département et le bassin d'emploi sont des découpages homogènes pour les trois derniers indicateurs alors que pour le premier, c'est-à-dire le taux de chômage, seul le département reste un découpage homogène.

On peut constater que ces indicateurs restent pour le moins assez globaux, bien qu'ils traduisent la différenciation du fonctionnement des marchés d'une zone à l'autre. Encore faut-il noter que cette affirmation ne semble pas tranchée, puisque l'on peut avoir par exemple des taux chômage voisins entre des zones de dimensions différentes alors que « les modalités concrètes de fonctionnement du marché (les flux d'embauche, de mobilité..., les migrations inter-zones...) sont radicalement différentes ». De plus, les quatre indicateurs caractérisent les conditions d'emploi des individus et non les conditions locales de l'offre d'emploi ; ils ne peuvent donc permettre de mesurer la spécificité locale des marchés. Il reste que, cette seconde méthode, tout comme la première, privilégie les mécanismes globaux d'ajustements dans un cadre homogène et on ne peut mettre en relief les liaisons pouvant exister entre les marchés.

Il serait fastidieux de dresser de manière exhaustive les méthodes ayant prévalu à la détermination des marchés locaux, tant elles sont nombreuses. On peut en passant citer les méthodes fondées sur l'analyse des correspondances (voir J.A. HERAUD et M. MOUGEOT ou encore H. NADINIER) ou sur la théorie des graphes etc.... Les multiples exemples que nous avons présentés jusqu'ici nous permettent de mieux cerner le problème posé lorsque l'on associe un marché local à un espace limité. Cette approche occulte les liaisons pouvant exister entre les marchés (les rapports de l'entreprise à l'espace nous ont montré que la grande entreprise avait la faculté d'utiliser plusieurs espaces) ; elle permet d'élaborer un cadre à l'intérieur duquel sont valorisés les outils d'analyse du modèle traditionnel du marché du travail. Nous allons avant de continuer, faire une synthèse d'ensemble et envisager comment on pourrait progresser grâce à un outil qui soit à même de prendre en compte non seulement les effets de structure inhérents à chaque marché, mais également leurs interrelations spatiales.

2.3.2.4 éléments de synthèse et recherche d'un outil d'analyse approprié pour l'étude des structures du marché.

Au terme des principales analyses qui ont été menées dans les précédents paragraphes, il en découle les observations suivantes :

- la dimension spatiale du marché du travail, ou pour simplifier le marché local du travail, repose sur un espace purement exogène, souvent d'origine administrative dont les limites n'ont pas beaucoup de signification. En effet, elles sont imposées soit par la pratique, soit par l'histoire, soit par des découpages institutionnels. Par ailleurs, même quand ces limites sont obtenues grâce à des méthodes statistiques assez pertinentes, il subsiste des problèmes de seuil qui mettent en cause leur relative stabilité.

-le marché local apparaît ainsi comme un processus d'allocation des emplois et de la main-d'œuvre ayant certaines limites dans l'espace. Il s'agit alors d'une reproduction à une échelle plus réduite des mécanismes globaux, la démarche n'étant pas en mesure de prendre en compte les spécificités inhérentes à chaque espace. Certes, ces limites sont nécessaires pour la politique locale de l'emploi, mais nous conduisent à une impasse sur le plan analytique puisqu'elles ne mettent pas en lumière les structures dans lesquelles fonctionne le marché. En somme la notion de marché local ici « n'est autre que celle d'un lieu où sont confrontées une offre et une demande, plus ou moins spécifiées selon les hypothèses que l'on fait sur le modèle. Des données a priori, déterminées ailleurs, servent à préciser l'espace géographique sur lequel on recherche l'équilibre »[69].

Au-delà, la question qui nous vient à l'esprit est de savoir s'il est possible de considérer le marché comme un univers clos avec des mécanismes de fonctionnement totalement autonomes. Cette approche revient à juxtaposer des espaces homogènes (on définit l'espace par un critère d'homogénéité) sans se soucier à ce qui les génère ou les relie. En effet, « un marché local ne se définit pas par l'absence de mouvements qui l'intègrent dans des espaces plus larges » [70]. Il serait plus justifié de fonder l'analyse du marché du travail sur une structure économique envisagée non comme une structure spatiale homogène, mais comme un ensemble d'espaces liés. D'ailleurs les rapports des différents agents économiques à l'espace ne sont que le reflet des liens pouvant exister entre les espaces. On l'a vu notamment avec les entreprises qui, dans le cadre de la politique de gestion de la main-d'œuvre, utilisent plusieurs espaces et on ne peut ne pas souligner les liens que ceux-ci entretiennent entre eux. Certes, les agents économiques modifient par leur pratique l'espace, mais cela ne doit pas masquer le fait que celui-ci induit également des comportements : ce sont les structures à l'origine de ces comportements qui semblent indispensables pour la compréhension du marché du travail dans sa spécificité. Au total, les structures d'emploi, de production, de formation, etc.. façonnent l'espace, et servent de support à l'analyse du marché

du travail. La perspective n'est plus celle que nous avons vu avec le marché limité géographiquement. Ici, le marché est perçu comme un processus « d'affectation de la main-d'œuvre dans les emplois, à travers le système de relations qu'entretiennent les agents ; les caractéristiques des postes ou des individus n'apparaissent plus comme une données en soi »[71]. On est alors dans une approche «segmentationniste» où les différences qui apparaissent entre les marchés sont totalement endogènes au système économique. Sur le plan spatial, l'approche envisage les modalités d'affectation de la main-d'œuvre dans des emplois en fonction de la situation géographique de chacun des marchés. Il s'agit alors de repérer les caractéristiques se situant à la fois dans l'espace et le marché du travail. Dès lors, l'intégration de l'espace dans le fonctionnement du marché du travail apparaît comme une mise en relation de ces différentes caractéristiques.

Le problème étant ainsi d'appréhender les structures de chacun des marchés, d'une part, et de voir comment les différents espaces sont reliés entre eux, d'autre part, il est opportun de faire appel à une technique d'analyse qui soit à même non seulement de concilier ces éléments d'approche, mais aussi adapté à la nature des données que nous aurons à exploiter : les variables les caractérisant sont celles relatives au marché du travail. L'outil que nous mettrons à profit est le modèle Log-linéaire. Il est de nature multiplicative et se prête fort bien à l'analyse structurelle des tableaux de données où celles-ci sont saisies à l'échelle multidimensionnelle. Par cet outil, nous mettrons en relief le facteur spatial qui découle des caractéristiques propres au marché.

Les trois chapitres qui suivent sont donc consacrés à la présentation du cadre conceptuel du modèle Log-linéaire, sa formulation et ses prolongements récents. Au-delà, nous l'appliquerons à l'offre d'emploi pour caractériser les structures spatiales du marché.

NOTES

- [1] MALLET : Le marché local du travail op. cité p. 3.
- [2] Haut Comité de l'Environnement : Rapport sur l'amélioration de la qualité de l'emploi, 1978.
- [3] BOUDEVILLE J.R. : Les espaces économiques Encyclopédie T. 9.
- [4] BOUDEVILLE J.R. : L'espace et les pôles de croissances, PUF.
- [5] Rapport du comité Emploi et Travail, 7^{ème} Plan, p. 246.
- [6] PERROUX F. : Unités actives et mathématiques nouvelles, Dunod, Collection Finance et Economie Appliquée, 1975, p. 82
- [7] PERROUX F. : Les espaces économiques, Economie Appliquée, Jan-Mars, 1950
- [8] GAMBIER D. : L'intégration de l'espace dans le fonctionnement du marché du travail : la notion du marché local, in Marchés locaux du travail, Documentation Française, p. 49.
- [9] MACLOUF P., SARBIB J.L. : Op. cité.
- [10] GAMBIER D. : Op. cité p. 50.
- [11] LOEIZ L. : L'espace et la crise, Crise et Espace, Economica, 1984.
- [12] SMITH A. : Recherches sur la nature et les causes de la recherche des nations, Gallimard 1976.
- [13] HERAND J.A., MOUGEOT M. : Migrations alternantes et bassins d'emploi : l'exemple de l'Alsace, RERU n°2, 1978.
- [14] MALLET L. : Op. cité p.114.
- [15] MALLET L. : Op. cité p. 115.
- [16] SALLEZ A. : Division spatiale du travail, développement régionale polarisé et théorie de la localisation, RERU n° 1, 1983. p. 70.
- [17] PLANQUE B. : Disjonction fonctionnelle et développement local, RERU n° 1, 1980.
- [18] SALLEZ A. : Op. cité.
- [19] AYDALOT PH. : La division spatiale du travail, in Espace et Localisation, ECONOMICA 1983.
- [20] SALLEZ A. : Op. cité p. 77
- [21] MACLOUF P., SARBIB J.L. : Op. cité.
- [22] AYDALOT P. : Le rôle du travail dans les nouvelles stratégies de localisation, RERU n°2, 1979 p. 178.
- [23] AYDALOT PH. : Le rôle du travail ..., op. cité p. 181
- [24] CAMPAGNA E. : Différenciation des espaces locaux dans les politiques de main-d'œuvre des grandes entreprises : l'exemple d'Usinor.
- [25] AYDALOT PH. : Op. cité p. 182.
- [26] GAMBIER D. : Voir 8, p. 68.

- [27] GAMBIER D. : Voir 8 p. 59.
- [28] INSEE : Statistiques et Indicateurs économiques des régions françaises Collection R.
- [29] BELLOC B., MARCHAND. : L'emploi régional en France depuis 1967, Eco et Stat, Nov./Dec. 1986.
- [30] INSEE : Disparités et diversité des régions françaises à la veille du 9^{ème} Plan, Archives et Documents.
- [31] PLANQUE B. : Technologies nouvelles et réorganisation spatiale, p. 98.
- [32] PLANQUE B. : Technologies nouvelles..., op. cité.
- [33] AYDALOT PH. : Dynamique spatiale et développement inégal. *Economica* 76.
- [34] SALLEZ A. : Les effets locaux des établissements décentralisés. Colloque ASRDLF 1982
- [35] PLANQUE B. : Technologies nouvelles ..., op. cité p. 101.
- [36] LEGARREC J. : Discours de clôture, colloque stratégie locale pour l'emploi, CEE p. 12, Fev. 83.
- [37] GAMBIER D., VERNIERES M. : Le marché du travail. *Economica* p. 128.
- [38] LAGUMIER J.F. : Pourquoi un renouveau de l'interventionnisme économique local dans les années 70, RERU, n°2, 1982, p. 155
- [39] Actes du colloque «une stratégie pour l'emploi », Ministère de l'emploi, Fev. 83.
- [40] LEDRUT R. : Sociologie du travail, PUF, p. 213 ;
- [41] CAVALLIER G. : La planification régionale et locale, Planification française n°181, 1977 ;
- [42] BAUVIR L., JAUMOTTE : La planification régionale et l'aménagement du territoire en Belgique. Cahier d'Economie Politique n° 7, 1979.
- [43] NEGRE M. : Modèles macroéconomique interrégionaux et seuils de complexité, RERU n°2, 1981.
- [44] Voir COURBIS R. : Modèles régionaux et modèles régionaux-nationaux. Collection Gama n°1, 1979.
- [45] Voir COURBIS R. : Analyse et planification nationale : le projet du modèle REGINA. Collection de l'INSEE, Série R n° 12 année 1973.
- [46] COURBIS R., PRAGER / Op. cité p. 20
- [47] PONSARD C ; : économie et espace, SESES, PARIS, 1955, p. 3
- [48] BENHAYOUN G. : L'emploi et le chômage des jeunes. Analyse urbaine.
- [49] JAYET H. : Les disparités régionales, ECO & Stat n° 153
- [50] SARTIN P. : Jeunes au travail, jeunes sans travail, Ed. d'Organisation p. 15.
- [51] MALLET L. : Op. cité p. 131.
- [52] SUDGEN R. : an application of search theory on the analysis of regional labor markets. *Reg. Sc. And Urban Econ.* 1980.
- [53] LEGAILLON J. : Les salaires . Ed. Cujas 1973. p. 65.

- [54] MALLET L. : Op. cité p. 132.
- [55] MICHEAU M. : Marché du travail local et Bassin d'emploi : état des études et évolution de la recherche, in Marchés locaux du travail p. 101-118.
- [56] CEE : Mission d'étude pour l'aménagement de la basse Vallée de la Seine. Fonctionnement des marchés locaux de l'emploi 1978.
- [57] MONNOYER M.C. : Détermination des zones vulnérables dans un espace industriel régional : application à la région Rhône-Alpes. RERU n°3. 1982.
- [58] MADINIER H. : Analyse de l'emploi salarié par zones. Colloque AEA-ASRDLF, Délégation de l'emploi.
- [59] DESTEPHANIS et al : Le fonctionnement d'un marché du travail local. Cahiers CEE n° 5, p. 18.
- [60] DESTEPHANIS et all : Op., cité, p. 9.
- [61] DESTEPHANIS et all : Op. cité
- [62] MALLET L. : Marché de la force du travail et localisation, Economie et planification, Documentation Française.
- [63] GAMBIER D., VERNIERES M. : Op. cité p. 133.
- [64] FLATRES : Voir MICHEAU M., « une notion récente d'aménagement du territoire : le bassin d'emploi, p. 93
- [65] HOUEE : Voir MICHEAU M., op. cité, p. 96
- [66] FRIEDMAN, DOUGLAS : Agropolitan development towards a new strategy for regional planning in Asia. UCLA papers Papers 1975.
- [67] Données sociales 197 INSEE.
- [68] TERRIER : Mirabelle, Courrier des Statistiques, n° 1
- [69] GAMBIER D : op. cité, p. 69.
- [70] MALLET op. cité, p. 218
- [71] Ici les caractéristiques des postes ou des individus découlent des mouvements dans lesquels ils s'insèrent et on peut selon l'expression de J.F. ROYER dire que le marché local est cette «infrastructure de mobilité » qui conduit les voyageurs dans «certaines stations ».

CHAPITRE III

CONCEPTS DE BASE NECESSAIRES A L'ANALYSE DU MODELE LOG-LINEAIRE

3.0 INTRODUCTION

L'étude du modèle Log-linéaire est au cœur de multiples investigations que nous esquisserons dans les chapitres ultérieurs. Bien que d'une très grande richesse (il permet de mettre en évidence les différents profils d'un tableau portant sur de très grands ensembles de données) ce modèle est encore d'usage peu courant en France. Afin de faciliter sa présentation, et la compréhension de nombreuses analyses qui en découlent, nous pensons par souci de commodité, qu'il est indispensable de définir au préalable les différents concepts lui servant de support et avec lesquels on se familiarisera par la suite.

Il s'agit avant tout d'un modèle descriptif, qui permet de décomposer chaque observation d'un tableau en un produit "d'effets" qui s'appliquent à cette observation, ce qui lui donne un caractère multiplicatif. En ce sens, on peut dire que le modèle Log-linéaire est un modèle descriptif multiplicatif. Nous y reviendrons plus en détail au chapitre IV.

Le modèle descriptif multiplicatif est un modèle s'adaptant très parfaitement aux variables qualitatives. Ce premier chapitre s'ouvre donc sur un rappel sur les variables (Cf. 3.1), d'où il ressort que l'application aux variables qualitatives des méthodes d'inférences statistiques liées aux variables quantitatives entraîne certaines incohérences.

D'autre part, ce modèle est utilisé pour l'étude des tableaux dits de contingence. La notion de tableau de contingence est définie au paragraphe 3.2, l'accent étant mis sur son approche multidimensionnelle et d'autres concepts qui en sont issus (totaux marginaux, marges). De plus, les observations à l'intérieur des cellules de ce tableau sont décrites par un certain nombre de processus d'échantillonnage (3.3) et nous verrons que ces processus appartiennent à la famille exponentielle (Cf. 3.3.4), dont le modèle multiplicatif est l'une des composantes.

Le paragraphe 3.4. aborde le principe du maximum de vraisemblance, sur lequel se fonde en général la procédure de "lissage" propre au modèle log-linéaire, c'est-à-dire la procédure d'estimation des observations d'un tableau pour éliminer les "bruits" dus aux fluctuations d'échantillonnage. Ce principe sera présenté ici dans sa généralité.

Quelques modèles - linéaires en général - sont enfin présentés au paragraphe 3.5. L'intérêt de cette présentation est de pouvoir mieux situer le modèle Log-linéaire.

3.1 RAPPEL SUR LES VARIABLES

Dans ce paragraphe, nous rappèlerons brièvement quelques éléments d'ordre général sur les variables permettant en même temps de distinguer les variables quantitatives des variables qualitatives. De surcroît, nous mettrons ensuite l'accent sur ces dernières, insistant par la même occasion sur la nécessité de leur adapter des outils spécifiques. La nature des données (elles sont qualitatives) sur lesquelles porteront nos recherches plus loin, justifiera également l'importance accordée aux variables qualitatives.

3.1.1 Eléments généraux.

En statistique, les données auxquelles nous sommes constamment confrontés sont souvent relatives à un ensemble bien déterminé qui est par exemple une population quelconque. Cet ensemble est repéré par un certain nombre de caractéristiques ou attributs, qui sont des variables.

On distingue en général deux groupes de variables : les variables quantitatives et les variables qualitatives. Les valeurs prises par les premières appartiennent à l'ensemble \mathbb{R} ou \mathbb{R}' (production journalière du pétrole, revenu) alors que celles prises par les secondes sont définies dans un ensemble moins structuré (région, sexe, etc.), et également caractérisées par leur forme discontinue, ce qui n'est pas le cas des variables quantitatives.

A l'intérieure de chacun des groupes, se retrouvent un certain nombre d'autres variables, les variables exogènes ou explicatives (en économie ce sont des variables "instrumentales " ou variables de politique économique), et les variables endogènes ou variables expliquées. Celles-ci sont en général expliquées par les variables exogènes dans un modèle de régression linéaire classique qui met en relief un schéma de causalité entre une variable endogène et une ou plusieurs variable explicatives. Les variables qualitatives constitueront ici notre principal pôle d'attraction, les variables quantitatives ayant déjà été au centre de plusieurs analyses statistiques. L'intérêt que nous porterons aux variables qualitatives est double :

- tout d'abord, la plupart des données statistiques issues des enquêtes, sondages, recensements, etc.. ont des caractères qualitatifs. Il peut s'agir des caractéristiques liées au sexe, au statut matrimonial, aux catégories socioprofessionnelles, etc.. pour ne citer que celles-là. L'ampleur prise par l'existence des données de telle nature mérite que l'on s'y attarde longuement afin d'entreprendre une étude plus approfondie

-ensuite, le développement des techniques de traitement adaptées aux variables qualitatives a connu un essor remarquable, surtout durant les deux dernières décennies. En témoigne d'ailleurs le volume des travaux qu'ont entrepris un certain nombre d'auteurs (Bishop, Fienberg et Holland (1975), Goodman (1979, 1981), Andersen (1980), Clifford Lunneborg (1983). Ces techniques ont donc la faculté majeure de tenir compte de la nature discontinue des variables qualitatives dont nous recensons ici les traits principaux.

3.1.2 Variables qualitatives

Nous évoquerons quelques caractéristiques des variables qualitatives, avant d'envisager la recherche d'une méthodologie appropriée pour leur traitement.

3.1.2.1 Caractéristiques des variables qualitatives.

Les variables qualitatives sont très couramment utilisées de nos jours dans les études empiriques (enquêtes sondages, etc..). En effet, l'exploitation des données sous forme catégorielle permet de mieux les traiter et d'en tirer le maximum d'information. Comme le dira Shizuhiko Nashizoto (1980) «l'emploi des catégories facilite généralement le travail et nous aide à retrouver l'information de façon pratique, par conséquent leur utilisation est devenue assez courante dans plusieurs branches». Une variable qualitative est dotée d'un certain nombre de modalités qui la définissent. Ainsi, la variable "sexe" a deux modalités : homme, femme.

Lorsque la variable qualitative a deux modalités (respectivement trois modalités) on dit qu'elle est dichotomique (respectivement trichotomique). En général, on dira qu'elle est polytomique si elle a plusieurs modalités.

Trois grands types de variables qualitatives sont imaginables.

- Les variables qualitatives nominales ou "naturelles". Ce sont des variables par essence même catégorielles. Ses modalités ne sont dotées d'aucune structure. C'est le cas, bien évidemment des variables comme le statut matrimonial, les catégories socioprofessionnelles.
- Les variables qualitatives ordonnées : il y a ici une question d'ordre et il s'agit des variables dont les modalités sont dotées d'une structure d'ordre. Cet ordre peut être naturel (cas des modalités des variables comme le temps, l'âge), ou établi sur la base d'un critère d'ordre. Ceci laisse supposer que toute variable qualitative, même nominale peut être ordonnée. Nous laissons pour l'instant de côté ce problème, tout en espérant y revenir ultérieurement (Cf. chapitre 5) pour étoffer de manière consistante les investigations donnant lieu au traitement des variables à modalités ordonnées.
- Les variables qualitatives « constituées » : ici, les modalités sont fixées de façon subjective ou a priori sur la base d'un critère. Nous allons l'illustrer par un exemple.

Soit la variable "régions de France". Elle a 22 modalités qui sont les 22 grandes régions de France. Sur la base d'un critère comme la taille par exemple, on peut regrouper les 22 régions en 6 grands ensembles homogènes de régions. Ces 6 grands ensembles forment les modalités d'une nouvelle variable - la variable " constituée" - que nous pouvons appeler " regroupement des régions de France suivant la taille". Une constatation est fort frappante : la frontière qui sépare cette troisième catégorie de la première est moindre et floue. Prenons par exemple la variable "catégories socioprofessionnelles (CSP)". Elle est par essence qualitative, c'est-à-dire naturelle. Cependant, ses modalités dépendent de façon subjective de la nomenclature en quatre classes (ouvriers, employés, cadres, professions libérales ou en sept classes (personnel des professions scientifiques, techniques, libérales et assimilées, directeurs et cadres administratifs supérieurs ; personnel administratif et travailleurs assimilés : personnel commercial et vendeurs ; travailleurs spécialisés dans les services ; agriculteurs, éleveurs, forestiers, pêcheurs et chasseurs ; et enfin ouvriers et manœuvres non agricoles et conducteurs d'engins de transport). Ces différents découpages peuvent bien nous amener à mettre les catégories socioprofessionnelles à l'actif des variables "constituées".

Puisque l'accent est surtout mis sur des variables qualitatives, la question qui reste d'emblée posée est de savoir si elles méritent d'être traitées par les méthodes d'inférences propres aux variables quantitatives. En effet, l'utilisation de telles méthodes à ces variables ne peut que fournir une information approximative les concernant, dans la mesure où "elles ne tiennent pas compte de l'absence de continuité et souvent de l'ordre naturel entre les modalités

que peut prendre le caractère "qualitatif". On s'arrête un instant à l'utilisation de ces méthodes afin de dégager certaines incohérences et par conséquent envisager d'autres techniques de traitement appropriées.

3.1.2.2 Recherche d'une méthodologie pour le traitement des variables qualitatives.

En général, l'application de l'inférence statistique liée aux variables quantitatives à des variables qualitatives consiste à codifier les variables qualitatives. Par le biais d'un tel mécanisme, ces variables sont quantifiées et on peut parler d'une "représentation quantitative des variables qualitatives". Ensuite, elles sont donc traitées comme si elles étaient quantitatives. En opérant ainsi, on aboutit à certaines invraisemblances méritant d'être soulignées. Pour s'en apercevoir, revenons à la variable catégorie socioprofessionnelle (CSP), définie avec les quatre modalités suivantes : « ouvrier, employé, cadre et profession libérale ». Soit z une « représentation quantitative » de la CSP définie par :

$$z = 1 \text{ si CSP = "ouvrier"}$$

$$z = 2 \text{ si CSP = "employé"}$$

$$z = 3 \text{ si CSP = "cadre"}$$

$$z = 4 \text{ si CSP = "profession libérale"}$$

Prenons deux grandeurs quelconques de z , soit 2 et 4. Leur moyenne est 3, ce qui correspond qualitativement à la catégorie "cadre", bien évidemment pour la codification utilisée (il est clair qu'il existe une infinité possible de codifications et que dans tous les cas la moyenne ne donne pas toujours 3, ce qui pose déjà le problème d'unicité des résultats). Par ailleurs, la grandeur 3 a quantitativement un sens, mais quand on regarde la correspondance avec la codification effectuée, il est invraisemblable que la catégorie "cadre" soit une catégorie moyenne des catégories "employé" et "profession libérale". D'ailleurs calculer une moyenne entre des catégories d'une variable qualitative n'a aucun sens. Il est toutefois possible de rechercher un profil moyen des différentes modalités ce qui est beaucoup plus conforme à la réalité.

Désignons par y une variable quantitative. La mesure de la liaison entre y et z peut se faire par le calcul des coefficients de corrélation. Or un tel calcul dépend de la codification adoptée pour z . Nous pouvons multiplier des exemples pour montrer l'inadaptation des

méthodes d'inférences liées aux variables quantitatives que l'on applique aux variables qualitatives. Il devient dès lors nécessaire d'appliquer à ces dernières les outils qui leur soient adaptés. Cependant, l'utilisation de ces outils requiert la mise en place d'un certain nombre d'éléments afin de mieux appréhender les analyses à entreprendre. Notre ambition, dans un cadre purement descriptif, est de mener les analyses à l'échelle multidimensionnelle. Nous dépassons donc le cadre bidimensionnel, c'est-à-dire le cadre des tableaux à double entrée, très souvent utilisés pour l'étude des données empiriques. Les résultats issus de tels tableaux sont toutefois globaux et manquent certaines informations supplémentaires que l'on tirerait, surtout dans le cadre des enquêtes par exemple, si l'on travaillait sur un tableau de dimension plus grande, permettant un affinement des résultats.

A titre illustratif, soit le tableau de dimension 2 ci-après, qui croise la population salariée française suivant les régions et les catégories socioprofessionnelles.

Tableau 4 : Croisement des régions et des catégories socioprofessionnelles.

CSP	1	2	i_2	I_2
REGIONS						
1						
2						
...						
i_1				$t(i_1, i_2)$		
...						
I_1						

$t(i_1, i_2)$ est l'effectif de la population salariée qui travaille dans la région i_1 , et qui appartient à la catégorie i_2 . Si à partir de ce tableau on peut dire que 10 % de personnes salariées travaillent dans la région i_1 et appartiennent à la catégorie socioprofessionnelle i_2 , on se rend tout de même compte que cette information est très globalisante. En effet, parmi ces 10 %, il y a une proportion d'hommes et de femmes qu'il faut relater, une répartition de ces salariés suivant l'âge ou toute autre variable étant à prendre en compte. Ainsi, au-delà des deux variables retenues pour le tableau 1, on peut ajouter deux autres variables (sexe, âge) pour

mieux affiner les résultats susceptibles d'être publiés, ce qui justifie l'introduction des tableaux multidimensionnels.

3.2. TABLEAU DE CONTINGENCE MULTIDIMENSIONNEL

Depuis les travaux de Pearson (1900) et de Yule (1900), les tableaux de contingence ont suscité un très grand intérêt, surtout dès 1935 avec l'article de Barlett, qui quitte le cadre traditionnel d'un tableau de dimension 2, pour traiter les variables catégorielles dans un tableau de dimension 3. Cet intérêt reste encore manifeste de nos jours.

Pour Pearson, un tableau de contingence n'est rien d'autre qu'une partition des distributions continues multivariées qui suivent la loi normale. Son souci majeur est celui de la recherche des liaisons pouvant exister entre les cellules d'un tableau, d'où l'évaluation du coefficient de corrélation des quatre cellules qui composent son tableau 2x2.

Yule, quant à lui, ne suppose pas continues les catégories d'un tableau quelconque. Il pense qu'elles sont fixées. A l'inverse de Pearson, il recherche une structure relationnelle entre les variables dans le tableau 2x2. C'est de ce vieux débat, comme le dira Fienberg (1978), qu'on peut mieux saisir de nos jours la notion de tableau de contingence.

Nous aurons ici à définir :

- le tableau de contingence,
- les totaux marginaux et marges,
- les marges impliquées et impliquantes.

3.2.1 Définition d'un tableau de contingence

On peut définir un tableau de contingence² comme un tableau qui donne la ventilation d'une population ou d'un échantillon aléatoire selon plusieurs caractères qualitatifs que l'on croise ou que l'on classe. Un tableau de contingence a deux caractéristiques dimensionnelles :

- la dimension proprement dite du tableau, qui est égale au nombre de variables que l'on croise dans le tableau ;

- le nombre de degrés de contraintes, c'est-à-dire le nombre de cellules élémentaires qui composent le tableau. C'est d'ailleurs à l'intérieur de ces cellules que sont rangés les individus, ceux-ci étant issus, comme nous l'avons souligné plus haut, d'une population ou d'un échantillon, et repérés par un ensemble de variables. Il est donc évident que toute construction d'un tableau de contingence passe par la définition de chacune des cellules de ce tableau. On peut donc construire le tableau de dimension 3 (tableau 2) ci-après :

Tableau 5 : Exemple d'un tableau à trois dimensions.

v_1	1			2			3		
v_3	1	2	3	1	2	3	1	2	3
v_2									
1			t(1,1,3)	t(2,1,1)		t(2,1,3)			
2									
3				t(2,3,1)					t(3,3,3)

Il s'agit d'un tableau 3x3x3 et qui a 27 cellules: $t(i_1, i_2, i_3)$ ($i_1=1,2,3$; $i_2=1,2,3$; $i_3=1,2,3$) est l'observation d'une cellule élémentaire (i_1, i_2, i_3) . Toutes les variables de ce tableau sont trichotomiques.

D'une manière générale, soit T un tableau dans lequel p variables V_1, V_2, \dots, V_p sont croisées et ayant respectivement I_1, I_2, \dots, I_p modalités. A chaque cellule, on a p-uples de modalités (une modalité par variable). Ce tableau est de dimension p et possède I degrés de contraintes:

$$I = \prod_{j=1}^p I_j$$

$t(i_1, i_2, \dots, i_p)$ est l'observation d'une cellule élémentaire avec :

$$i_1 = 1, \dots, I_1$$

$$i_2 = 1, \dots, I_2$$

.

.

$$i_p = 1, \dots, I_p$$

L'une des caractéristiques dominantes d'un tableau de contingence, c'est la possibilité d'y effectuer des sommations pour obtenir des grandeurs ayant un sens concret. Ces sommations permettent de définir les notions de totaux marginaux et marges qui auront une importance dans la modélisation log-linéaire.

3.2.2 Totaux marginaux et marges.

Soit T un tableau de contingence multidimensionnel à I cellules. Ce tableau est bien évidemment de dimension p avec comme cellule élémentaire (i_1, i_2, \dots, i_p) . Des sommations partielles qui portent sur une ou plusieurs modalités de variables seront effectuées dans ce tableau. En procédant ainsi on obtient des totaux marginaux, ceux-ci étant définis par les modalités qui n'ont pas été retenues dans la sommation

Fixons-nous une modalité quelconque de la $2^{\text{ème}}$ variable. On définit $t(+, i_2, +, +)$ par :

$$t(+, i_2, +, \dots, +) = \sum_{i_1=1}^{I_1} \sum_{i_3=1}^{I_3} \dots \sum_{i_p=1}^{I_p} t(i_1, i_2, \dots, i_p) \quad (\text{E.3.1})$$

où $t(+, i_2, +, \dots, +)$ est un total marginal. Il est de dimension 1 et on dit dans ce cas qu'il se situe dans la cellule marginale i_2 (i_2 étant fixé bien évidemment.). Cependant i_2 varie ($i_2 = 1, \dots, I_2$). On forme alors l'ensemble des totaux marginaux d'ordre 1 correspondant aux modalités de la $2^{\text{ème}}$ variable. C'est cet ensemble de totaux marginaux qui constituent la marge notée $M(i_2)$ et telle que :

$$M(i_2) = \{t(+, i_2, +, \dots, +); i_2 = 1, 2, \dots, I_2\} \quad (\text{E 3.2})$$

$M(i_2)$ a I_2 degrés de contraintes.

Concrètement, si on revient au tableau 1, $t(i_1, +)$ qui est un total marginal représente l'effectif des salariés d'une région i_1 quelconque, alors que la marge correspondante est le vecteur dont les composantes sont les effectifs pour chaque région.

De manière analogue à (E 1.1), s'établissent aussi les totaux marginaux d'ordre 2 ou d'ordre supérieur à 2, ainsi que les marges associées.

$$t(+, i_2, i_3, +, \dots, +) = \sum_{i_2=1}^{I_2} \sum_{i_3=1}^{I_3} \dots \sum_{i_p=1}^{I_p} t(i_1, i_2, \dots, i_p) \quad (E.3.3)$$

est un total marginal d'ordre 2 se situant dans la cellule élémentaire (i_2, i_3) .

$M(i_2, i_3)$ est la marge associée et définie par :

$$M(i_2, i_3) = \{t(+, i_2, i_3, +, \dots, +) : i_2=1, 2, \dots, I_2\} \quad (E.3.4)$$

Bien évidemment $M(i_2, i_3)$ a $I_2 * I_3$ degrés de contraintes.

Plus généralement, soient i_ω et i_Ω deux indices unidimensionnels ou multidimensionnels d'ordre q et s respectivement ($q \leq p$, $s \leq p$ et $q + s = p$)

On a donc $i_\omega \cup i_\Omega = (i_1, i_2, \dots, i_p)$ (\cup est le symbole de la réunion des ensembles).

$$t(+, +, \dots, +, i_\omega, +, \dots, +) = \sum_{\Omega} t(i_1, i_2, \dots, i_p) \quad (E.3.5)$$

est le total marginal d'ordre q se situant dans la cellule élémentaire (i_ω)

$$m(i_\omega) = \{t(+, \dots, +, i_\omega, +, \dots); i_\omega = 1, \dots, I_\omega / \omega = 1, \dots, q\} \quad (E.3.6)$$

Cette notion de marge doit être complétée par celle des marges impliquantes et des marges impliquées.

3.2.3 Marges impliquantes et marges impliquées.

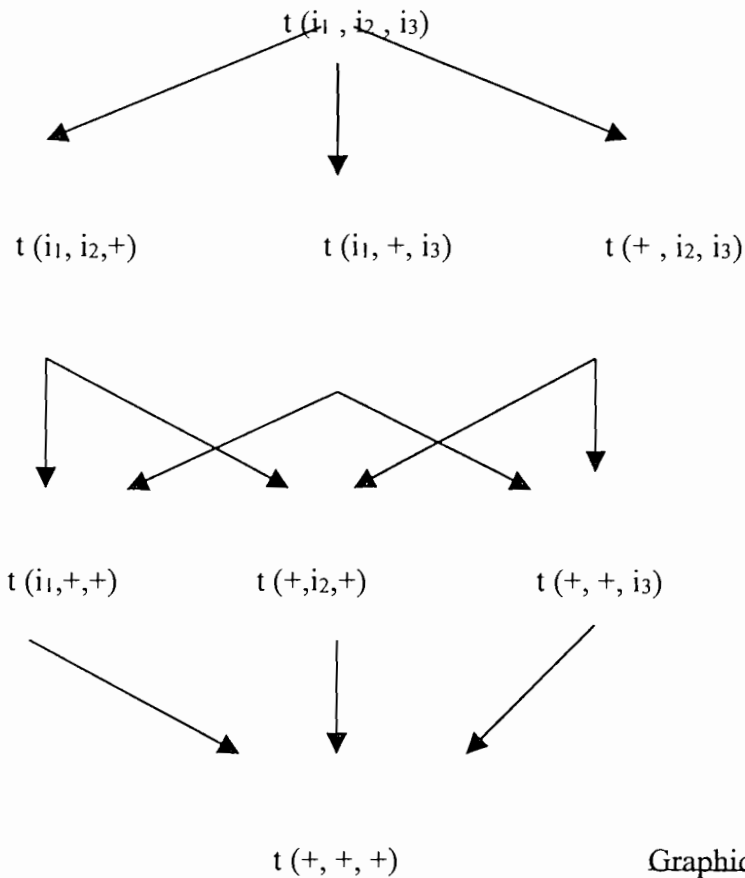
Les marges impliquées s'obtiennent à partir des équations (E.3.2.), (E.3.4), (E.3.6). En effet, lorsque l'on part des marges d'ordre q , on peut définir les autres marges d'ordre s ($s < q$) telles que le groupe d'indices représentant la marge d'ordre s soit inclu dans le groupe d'indices représentant la marge d'ordre q . La marge d'ordre q est appelée marge impliquante alors que la marge d'ordre s est impliquée. Formellement, on a :

$$B_1 = \{i_\omega / \omega = 1, \dots, q\} \quad B_2 = \{i_\alpha / \alpha = 1, \dots, s\} \quad B_2 \subset B_1$$

$$i_j \in B_1$$

$$\sum_{i_j} t(+, \dots, i_\omega, \dots, +) = t(+, +, i_\alpha, \dots, +) \quad (E.3.7)$$

De l'équation (E 3.7), le terme de gauche représente la marge impliquante et celui de droite la marge impliquée. On constate donc que toutes les marges sont impliquées par la marge de dimension p qui est le tableau T . Si on prend le tableau 2, on peut voir cette relation d'implication grâce au graphique suivant :



Graphique 1

Ce graphique met en lumière l'implication de toutes les marges du tableau par la marge d'ordre 3 $t(i_1, i_2, i_3)$ qui est la marge impliquante. De même $t(+, +, +)$ qui est la marge d'ordre 0 est impliquée par toutes les autres marges. Le sens des flèches indique l'ordre d'implication.

Toutes ces notions de totaux marginaux, marge, marge impliquante et impliquée s'appliquent à des tableaux de contingence. Peut-on également les concevoir dans d'autres types de tableaux et plus particulièrement les tableaux non de contingence ? En effet, appliquées à ces derniers, ces notions seront dépourvues de sens dans la mesure où les tableaux non de contingence sont constitués d'un ensemble de variables hétérogènes. Cependant, il existe des transformations pouvant conduire un tableau non de contingence à un tableau de contingence.

On peut opérer une des transformations en se servant du tableau suivant qui croise en volume la production des biens de consommation (blé, oignons, poires, tomates) suivant des régions françaises Bretagne, Centre, Picardie, Nord

Tableau 6 : Exemple d'un tableau qui n'est pas de contingence.

Biens \ Régions	Bretagne	Centre	Picardie	Nord
Blé				
Oignon		$x(i_1, i_2)$		
Poires				
Tomates				

Pour toute cellule élémentaire (i_1, i_2) ($i_1=1, \dots, 4$; $i_2=1, \dots, 4$), $x(i_1, i_2)$ représente donc le volume du bien i_1 produit dans la région i_2 .

Bien que le total marginal $x(i_1, +)$ (c'est la quantité du bien i_1 produite dans toutes les régions) ait un sens, ce tableau n'est pas de contingence. Certes, mathématiquement, on peut calculer $x(+, i_2)$ qui représente le volume de tous les biens dans une région i_2 . Mais quelle signification peut avoir cette grandeur ? Peut-on sommer les poires et les tomates ? Toute sommation ne peut avoir un sens que si elle porte sur des grandeurs homogènes. En connaissant le système de prix des différents biens, la pondération des grandeurs du tableau 3 permet de les exprimer plutôt en valeur. Cette homogénéisation rend possible tout passage d'un tableau qui n'est pas de contingence à un tableau dit de contingence. Même si une telle transformation ne s'impose pas dans les tableaux à exploiter dans nos travaux, il valait la peine qu'on l'illustrât, ce d'autant plus que la plupart des tableaux issus des enquêtes ou sondages sont caractérisés par leur "conditionnement" assez mauvais, qui est la conséquence de la coexistence d'une multitude de variables de nature différente. Ce qui exige au moins, avant d'entreprendre toute analyse, de les rendre homogène. Il faut également signaler que l'outil utilisé pour l'exploitation d'un tableau reste tributaire de la nature de ce tableau. C'est pourquoi le modèle log-linéaire ne peut être appliqué qu'aux tableaux de contingence. D'autre

part, les observations à l'intérieur de ces tableaux peuvent être décrites par des processus d'échantillonnage que sont les processus poissonnien, multinomial et produit-multinomial, couramment utilisés dans l'approche Log-linéaire.

3.3. QUELQUES PROCESSUS D'ECHANTILLONAGE

On conçoit généralement trois processus d'échantillonnage pour le traitement des tableaux de contingence, du moins dans l'approche multiplicative. Afin de cerner les caractéristiques de chacun, nous présentons dans les sous-sections qui suivent:

- le processus d'échantillonnage poissonnien,
- le processus d'échantillonnage multinomial,
- le processus d'échantillonnage produit-multinomial.

3.3.1 Processus d'échantillonnage poissonnien.

Un processus de ce type suppose qu'il n'existe aucune contrainte en ce qui concerne les marges du tableau de contingence. Cependant, il est intéressant, avant de le formuler, de préciser comment un phénomène peut être considéré comme Poissonnien.

Soit un intervalle de temps fixe (γ_1, γ_2) . Si la probabilité qu'un événement se produise entre γ et $\gamma + \partial\gamma$ ne dépend que de $\partial\gamma$, alors le processus est poissonnien.

Désignons par T un tableau de contingence de dimension 3 (Cf. tableau 2) où l'observation dans une cellule élémentaire est $t(i_1, i_2, i_3)$. L'espérance d'une cellule quelconque est $m(i_1, i_2, i_3)$, c'est-à-dire $E[t(i_1, i_2, i_3)] = m(i_1, i_2, i_3)$.

Le tableau T a N observations et on suppose qu'elles sont toutes réparties indépendamment sur les $I_1 * I_2 * I_3$ cellules $(i_1=1, \dots, I_1; i_2=1, \dots, I_2; i_3=1, \dots, I_3)$.

Si chaque observation suit un processus poissonnien, elle peut être décrite par la fonction de densité suivante :

$$\frac{m(i_1, i_2, i_3)^{t(i_1, i_2, i_3)} e^{-m(i_1, i_2, i_3)}}{t(i_1, i_2, i_3)!}$$

Les observations étant supposées indépendantes, on obtient pour toutes les cellules du tableau la fonction de vraisemblance ci-après :

$$\prod_{i_1, i_2, i_3} \frac{m(i_1, i_2, i_3)^{I(i_1, i_2, i_3)} e^{-m(i_1, i_2, i_3)}}{I(i_1, i_2, i_3)!} \quad (\text{E 3.9a})$$

Fisher fut l'un des premiers à utiliser la fonction (E3.9a) pour l'étude des tableaux de contingence. Elle est largement exploitée aujourd'hui par les auteurs qui s'intéressent aux modèles multiplicatifs, notamment les modèles log-linéaires. On pourra ajouter à ce processus un autre processus d'échantillonnage : le processus multinomial.

3.3.2. Processus d'échantillonnage multinomial

Le processus précédent est formulé sans aucune contrainte de marge du tableau T. Cependant, dans le cas du processus multinomial, une contrainte est à retenir : celle qui correspond au grand total N (marge d'ordre 0). Dès lors, le total qui était variable dans le processus poissonnien devient fixe. Cette contrainte est telle que la somme de toutes les observations du tableau soit égale à N, d'où :

$$\sum_{i_1} \sum_{i_2} \sum_{i_3} I(i_1, i_2, i_3) = N \quad (\text{E 3.10})$$

Plus explicitement, T étant un tableau de contingence de dimension 3, on note $p(i_1, i_2, i_3)$, la probabilité qu'une observation quelconque tombe dans la cellule élémentaire (i_1, i_2, i_3) , ce qui revient à:

$$\begin{cases} p(i_1, i_2, i_3) > 0 \\ \sum \sum \sum p(i_1, i_2, i_3) = 1 \end{cases} \quad (\text{E 3.11})$$

Puisque les observations suivent toutes un processus multinomial, la vraisemblance du tableau est :

$$\frac{N!}{\prod t(i_1, i_2, i_3)} \prod [p(i_1, i_2, i_3)]^{t(i_1, i_2, i_3)}$$

$$E(t(i_1, i_2, i_3)) = N p(i_1, i_2, i_3) \quad (\text{E 3.12a})$$

Remarque 1 :

La fonction (E 3.9a) conditionnelle donne (E 3.12a). En d'autres termes, si nous prenons la fonction (E 3.9a) sachant que $t(+,+,+) = N$, alors on retrouve la fonction (E 3.12a) (cf. Andersen, page 149). Les deux processus précédents sont couramment utilisés pour le lissage log-linéaire. Néanmoins, il en existe un troisième, peu courant, qui est le processus produit-multinomial.

3.3.3. Processus d'échantillonnage produit-multinomial

Dans un tableau de contingence multidimensionnel, le processus multinomial a pour caractéristique fondamentale de prendre en compte, en ce qui concerne les marges, uniquement celles correspondant au grand total. Toutefois, il existe des situations où l'on peut considérer comme fixes certains totaux marginaux. Ces situations interviennent quand une ou plusieurs variables du tableau multidimensionnel sont explicatives et les autres expliquées. Dans ce cas précis, le modèle produit-multinomial permet de mieux décrire les observations à l'intérieur de chaque cellule. Les marges correspondant aux variables explicatives sont fixées et on effectue leur classification suivant les variables expliquées, ce qui donne un processus produit-multinomial pour la ou (les) marge(s) considérée(s). De manière formelle, on a pour un tableau tridimensionnel T où la marge $M(i_1)$ est supposée fixée (c'est-à-dire tous les totaux $t(i_1, +, +)$, $i_1 = 1, \dots, I_1$).

$$p(i_1, i_2, i_3) = \frac{p(i_1, i_2, i_3)}{p(i_1, +, +)}$$
 est la probabilité conditionnelle qu'une observation quelconque

prenne les modalités i_2 et i_3 des variables 2 et 3 sachant qu'elle prend la modalité i_1 de la première variable ($i_1 = 1, \dots, I_1 ; i_2 = 1, \dots, I_2 ; i_3 = 1, \dots, I_3$). En adoptant ce processus on obtient la vraisemblance des observations du tableau :

$$\prod_{i,j} \frac{t(i,j,i)!}{\prod_{i,j} t(i,j,i)!} \prod [p'(i,j,i)]^{(0,1,0)}$$

Remarque 2 :

La fonction (E.3.12) conditionnelle, c'est-à-dire sachant $t(i_1,+,+) = n(i_1,+,+)$ donne la fonction (E 3.13 a) (voir, Andersen, page 150). D'une manière générale, soit i_ω un indice unidimensionnel ou multidimensionnel et $t(i_\omega)$ la valeur observée à l'intérieur d'une cellule élémentaire i_ω et $m(i_\omega)$ l'espérance de $t(i_\omega)$ ($E(t(i_\omega)) = m(i_\omega)$). Les fonctions (E 3.9 a), (E3.12a), (E 3. 13 a) s'écrivent :

$$\prod_{i,j,i} \frac{m(i_1,i_2,i_3)^{t(i_1,i_2,i_3)} e^{-i,j,i}}{t(i,j,i)!}$$

Les estimateurs du maximum de vraisemblance (cf. 3.4) des trois processus sont les mêmes. Cette identité est largement reprise par Birch (1963), Haberman (1974), Bishop et Holland (1973), Clifford E. Lunneborg (1933), etc..

Au-delà de la définition des tableaux de contingence (Cf. 3.2.1), nous venons de voir les processus sous quelques différentes formes qui permettent de saisir les observations à l'intérieur des cellules de ces tableaux. Ces processus ont une caractéristique particulière : ils font partie d'une famille de modèles qu'on appelle famille exponentielle. Celle-ci mérite qu'on y fasse un détour, ce d'autant plus que la plupart des statistiques inhérentes au modèle log-linéaire s'expriment par son biais. On verra par exemple comment il en découle les statistiques suffisantes indispensables à toute procédure d'estimation (Cf.3.4).

3.3.4 Famille exponentielle.

La famille exponentielle ne suscite pas d'intérêt seulement de nos jours. Déjà, Lehmann (1959) l'avait étudiée dans le cadre d'une théorie pour les tests (test du rapport de maximum de vraisemblance). Bien avant lui, Fisher (1934) et Rao (1952) en faisaient une étude dans leurs écrits. Dans une approche purement Log-linéaire, la famille exponentielle a été

développée par Birch (1963), Goodman (1973), Plackett (1974), Barndorff-Nielsen (1978) et Andersen (1930). Ces différentes études permettent de mieux comprendre sa définition.

3.3.4.1 Définition d'une famille exponentielle.

Soient T un tableau de contingence p -dimensionnel et t le vecteur lexicographique représentant les N observations de ce tableau. On suppose que celles-ci sont indépendamment réparties sur les I cellules $\left(I = \prod_{j=1}^p I_j \right)$ du tableau.

Le tableau T sera considéré comme ayant I variables aléatoires $T(1), \dots, T(I)$ tels que $t(k)$ ($k = 1, \dots, I$) est la valeur prise par la variable aléatoire $T(k)$.

La probabilité d'observer une valeur du tableau est :

$$P(T(k)=t(k)) = f(t(k)/\alpha(1), \alpha(2), \dots, \alpha(q)) \quad (E 3.14)$$

où $\alpha(1), \dots, \alpha(q)$ est une ensemble de q paramètres inconnus.

La probabilité d'observer toutes les valeurs prises par $t(k)$ ($k = 1, \dots, I$) devient :

$$P[T(1)=t(1), T(2)=t(2), \dots, T(I)=t(I)] = \prod_{k=1}^I p[T(k) = t(k)] \quad (E 3.15)$$

du fait de l'indépendance des observations.

Compte tenu de (E 3.14), on a :

$$f(t(1), \dots, t(I)/\alpha(1), \dots, \alpha(q)) = \prod_{k=1}^I f(t(k)/\alpha(1), \dots, \alpha(q)) \quad (E 3.15b)$$

Cette fonction de densité définie, le problème est de savoir dans quelle mesure elle appartient à la famille exponentielle.

En effet, soit D une distribution quelconque et soit d la valeur prise par D . Alors D appartient à la famille exponentielle si la fonction de densité associée à d est telle que :

$$f[(d/\gamma(1), \dots, \gamma(q))] = v^{-1}[\gamma(1), \dots, \gamma(q)] e^{\sum_{g^{(d)} x, (\gamma(1), \dots, \gamma(q))}}$$

$$\text{avec } s_0(\gamma(1), \dots, \gamma(q)) = 1 \quad (\text{E 3.16})$$

u est la dimension de la famille exponentielle et c'est le plus petit nombre pour lequel (E3.14) peut se mettre sous la forme (E 3.16).

En revenant à la fonction de densité (E 3.5b) du tableau, son expression sous forme exponentielle devient.

$$f(t(1), \dots, t(I) / \gamma(1), \dots, \gamma(q)) = v^{-1}[\gamma(1), \dots, \gamma(q)] e^{\sum_{g, t^{(k)} x, (\gamma(1), \dots, \gamma(q))}} \quad (\text{E 3.17})$$

l'écriture de (E 3.17) sous forme canonique (Cf. Andersen (1980)) est ci-après:

$$\begin{aligned} f(t(1), \dots, t(I) / \alpha(1), \dots, \alpha(q)) &= v^{-1}[\alpha(1), \dots, \alpha(q)] e^{\sum_{g, t^{(k)} x, (\alpha(1), \dots, \alpha(q))}} \\ &= v^{-1}[\alpha(1), \dots, \alpha(q)] e^{\alpha^{(j)} \sum_{g, t^{(k)}}} e^{\sum_{g, t^{(k)}}} \end{aligned}$$

avec

$$\alpha(1) = s_1(\gamma(1), \dots, \gamma(q))$$

$$\alpha(u) = s_u(\gamma(1), \dots, \gamma(q))$$

La forme exponentielle précédente est d'un grand intérêt et celle-ci sera plus manifeste quand nous aborderons la procédure d'estimation plus spécifiquement dans le modèle log-linéaire. Cette procédure, qui est celle du principe du maximum de vraisemblance en général, requiert en amont la définition d'un ensemble de statistiques, notamment les statistiques suffisantes qui s'obtiennent plus facilement lorsque le modèle appartient à la famille exponentielle.

3.3.4.2 Statistiques suffisantes

Soit C une statistique. Elle est suffisante si elle peut fournir toute l'information relative à une population -un échantillon- donnée, cette information étant identique à celle que l'on obtiendrait par l'intermédiaire des paramètres inconnus.

Selon le critère de Neyman (1938), C est une statistique suffisante si :

$$f[t(1), \dots, t(I)/\alpha] = g(c/\alpha)h(t(1), \dots, t(I)) \quad (E 3.19)$$

En identifiant (E 3.18) à (E 3.19) on déduit que :

$$g(c/\alpha) = v^{-1}[a(1), \dots, a(u)] e^{\sum_{a(j)} \sum_{g_j(t(k))} c_j} \text{ et}$$

$$h(t(1), \dots, t(I)) = e^{\sum_{g_0(t(k))} c_0}$$

Posons : $C_j = \sum_k g_j(t(k))$. On tire $g(c/\alpha) = v^{-1}[\alpha(1), \dots, \alpha(u)] e^{\sum_{a(j)} c_j \alpha(j)}$.

Finalement on constate que $g(c/\alpha)$ dépend de $\alpha(j)$ et $c_j(j = 1, \dots, u)$ alors que $h(t(1), \dots, t(I))$ ne dépend que de C_j . La conclusion est que (C_1, \dots, C_u) est un ensemble de statistiques suffisantes pour le calcul des paramètres $\alpha(1), \alpha(2), \dots, \alpha(u)$.

On dira que C est minimale si elle s'exprime comme une fonction de toutes les autres statistiques suffisantes. Dans ce cas, elle est unique. Nous verrons comment rechercher les statistiques minimales suffisantes dans le cadre particulier du modèle log-linéaire (Cf. chap.4). Ces statistiques sont indispensables à la procédure d'estimation des observations d'un tableau. Certes, il existe plusieurs procédures d'estimation, mais nous choisissons de présenter pour l'instant celle du principe du maximum de vraisemblance sur laquelle seront fondées nos estimations ultérieures.

3.4. ESTIMATION DU PRINCIPE DU MAXIMUM DE VRAISEMBLANCE EN GENERAL

Le principe du maximum de vraisemblance est l'une des méthodes d'estimation les plus couramment utilisées en statistique. Cette méthode est évidemment fondée sur la fonction de vraisemblance qu'il faut optimiser ; Fisher (1925) a introduit ce principe pour l'estimation pendant que Neyman et Pearson (1938) l'exploitent pour l'élaboration des tests. C'est surtout Lehmann (1959) qui développe la théorie des tests fondée sur le principe du maximum de vraisemblance. Avant d'explicitier ce principe, il nous faut poser le problème qui lui est sous-

jacent. D'autre part, ce problème est celui d'optimisation sous contraintes d'où la nécessité de définir les types de contraintes à rencontrer -contraintes internes (3.4.2) et contraintes externes (3.4.3) -et susceptibles d'être imposées aux deux approches de la procédure-. l'approche directe et l'approche indirecte (3.4.4.2).

3.4.1 Problème d'estimation par le principe du maximum de vraisemblance.

Nous restons toujours dans la logique d'un tableau de contingence auquel on applique un modèle d'analyse (structure statistique). Ce modèle est caractérisé par un ensemble de paramètres à estimer. Si nous retenons le tableau défini en (3.3.4.1), la fonction de vraisemblance des observations de ce tableau est donnée par (E 1.18) où le vecteur $\alpha(1), \alpha(2), \dots, \alpha(u)$; $\alpha(1)$ appartient à un espace γ quelconque appelé espace des paramètres ; γ est un espace u -dimensionnel.

Le problème consiste à chercher dans r un vecteur α qui rende (E 3.18) maximum, ce qui se traduit formellement par :

Rechercher $\hat{\alpha}$ tel que

$$f(t(1), \dots, t(I) / \hat{\alpha} = \sup_{\alpha \in \gamma} f(t(1), t(2), \dots, t(I) / \alpha) \quad (\text{E 3.20})$$

Un autre problème analogue, permettant de rechercher le vecteur, est celui de la minimisation de l'information discriminante (cf. chap.4) qui n'est rien d'autre que le dual de (E 3.20). Dans tous les cas, il existe des contraintes à prendre en compte. Ce sont le plus souvent des contraintes de marges. A cet égard nous retiendrons les contraintes internes et les contraintes externes.

3.4.2 Contraintes internes

Les contraintes internes sont d'usage courant dans les problèmes d'estimation par le principe du maximum de vraisemblance. L'idée qui prévaut ici est la suivante : on cherche α

du problème (E 3.20) qui permet en même temps de "lisser" le tableau t tout en conservant ses marges.

Formellement on a t le vecteur lexicographique des observations à l'intérieur des cellules élémentaires du tableau t à I cellules : t est le vecteur $(I \times 1)$ et T la matrice de désignation ou « design matrix ». Elle décrit le nombre et le genre de contraintes à imposer au vecteur t^* qui est le vecteur estimé de t . T est une matrice $(I \times r)$ où r est le nombre de contraintes linéaires existantes. Si toutes les contraintes sont choisies de manière à ce que t^* possède certaines caractéristiques de t , les contraintes internes sont traduites par :

$$Tt = Tt^* \quad (E 3.21)$$

S'agissant du tableau (2×2) (cf. tableau 4) par exemple

Tableau 7 : Exemple d'un tableau 2×2 .

	1	2	Totaux
1	$t(1,1)$	$t(1,2)$	$t(1,+)$
2	$t(2,1)$	$t(2,2)$	$t(2,+)$
Totaux	$t(+,1)$	$t(+,2)$	$t(+,+)$

En se donnant une matrice de " désignation" qui respecte les totaux marginaux $t(+,1)$ et $t(+,2)$, alors T est définie par:

$$\begin{vmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{vmatrix} \text{ de sorte que}$$

$$Tt = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} t(1,1) \\ t(1,1) \\ t(1,1) \\ t(1,1) \end{bmatrix} = \begin{bmatrix} t(+,1) \\ t(+,2) \end{bmatrix} = \begin{bmatrix} t^*(+,1) \\ t^*(+,2) \end{bmatrix} = Tt^*$$

Les contraintes internes ne sont pas les seules, puisqu'on peut concevoir également des contraintes externes.

3.4-3. Contraintes externes

Deming et Stephan (1940) pensent que le lissage d'un tableau de contingence représentant une population ou un échantillon donné doit respecter certains totaux marginaux connus a priori. C'est ce qu'il est convenu d'appeler, selon Gokhale et Kullback (1978) contraintes externes. A la différence des contraintes internes, dont l'objet est de rechercher un vecteur t^* ressemblant le plus à t , on cherche ici t^* proche d'une structure idéale que l'on se donne. Lorsque les contraintes externes sont respectées, on obtient :

$$T't = \Phi \quad (\text{E } 3.22)$$

où Φ est le vecteur ($r \times 1$) des r contraintes externes.

Les contraintes externes peuvent intervenir dans les opérations de « recollage » des données où l'on se donne des marges a priori. De même, on peut regarder l'évolution de la structure d'un échantillon d'une population en prenant, pour marges celles issues des recensements.

Remarque : Les deux catégories de contraintes précédentes s'appliquent aux modèles Log-linéaire en général. Les matrices associées à ces contraintes sont composées uniquement des éléments 1 et 0 (existence ou non d'un "effet").

D'autres types de contraintes n'ont pas cette caractéristique. C'est le cas par exemple des contraintes que l'on rencontre dans les problèmes d'optimisation d'une fonction économique et qui traduisent des objectifs à atteindre ou à dépasser. Ainsi, un industriel voulant maximiser son profit se verra contraint par les quantités d'input (on ne peut pas utiliser indéfiniment les inputs pour les raisons de coût entrant dans le processus de production). De même, un ménage qui consomme un certain assortiment de biens maximisera son utilité de façon à ne pas dépasser ces possibilités financières (contraintes budgétaires). Nous ne nous étendons pas à la recherche d'autres formes de contraintes (il en existe une multitude). L'exemple que nous venons d'évoquer marque la différence existant entre les contraintes que nous utiliserons dans le modèle log-linéaire (cas des contraintes (E 3.21) ou (E 3.22) et qui reflètent une structure) et d'autres types de contraintes comme les précédentes qui expriment une limite.

Comme on l'a déjà mentionné plus haut, les contraintes (E 3.21) ou (E3.22) seront prises en compte dans la procédure d'estimation dont nous décrivons ici quelques méthodes.

3.4.4 Méthodes d'estimation en général

On conçoit assez souvent deux approches : l'approche directe et l'approche itérative, celle-ci s'effectuant quand la première ne peut être appliquée.

3.4.4.1 Méthode d'estimation directe

Pour illustrer l'approche directe, revenons au problème (E 3.20).

$$\underset{\alpha}{SUP} f(t(1), t(2), \dots, t(I) /) = \underset{\alpha}{SUP} v^{-1} [\alpha(1), \dots, \alpha(u)] e^{\sum \alpha(j)c_j} \\ h(c_1, c_2, \dots, c_u) \quad (E 3.23)$$

posons $L(\alpha) = f[t(1), \dots, t(I) / \alpha]$

$$\rho(\alpha) = \log L(\alpha) = -I \text{Log} v(\alpha(1), \alpha(2), \dots, \alpha(u)) + \sum \alpha(j)c_j + \text{Log} h(\alpha(1), \dots, \alpha(u)) \quad (E 3.24)$$

Maximiser $L(\alpha)$ revient à maximiser $\rho(\alpha)$

Les conditions de 1er ordre sont définies par :

$$\rho'(\alpha) = 0 \Rightarrow \frac{\partial \text{Log} L(\alpha)}{\partial \alpha(j)} \\ \forall j=1, \dots, u \\ - I \frac{\partial \text{Log} v(\alpha(1), \dots, \alpha(u))}{\partial \alpha(j)} + c_j = 0$$

$$\text{On tire enfin } c_j = \frac{I \partial \text{Log} v(\alpha(1), \dots, \alpha(u))}{\partial \alpha(j)}$$

$$\forall j=1, \dots, u \quad (E 3.25)$$

$\hat{\alpha}(1), \hat{\alpha}(2), \dots, \hat{\alpha}(u)$ est obtenu en résolvant simultanément les u équations de (E 3.25). On n'arrive pas tout le temps aux solutions du type (E3.25) dans la mesure où certaines fonctions de vraisemblance sont telles que $\rho'(a)$ n'est pas nulle. Le recours à la procédure itérative intervient alors dans ce cas.

3.4.4.2 Approche par la procédure itérative.

La procédure itérative s'appuie sur l'un des deux principaux algorithmes suivants pour le calcul des estimateurs de vraisemblance : l'algorithme itératif d'ajustement proportionnel (Fienberg (1970), Gokhale (1971), Haberman (1974) et l'algorithme de Newton-Raphson (Andersen, 1980). Historiquement, l'algorithme itératif d'ajustement proportionnel est la première méthode à être utilisée (Duncan, 1940), et il est caractérisé par une simplicité dans sa mise en œuvre. Néanmoins, il converge moins vite que l'algorithme de Newton-Raphson.

Nous décrirons l'algorithme de Newton-Raphson, juste pour donner une idée du mécanisme de la procédure itérative.

Soit donc $\rho(\alpha)$ telle que définie en (E 3.24). On suppose que $\rho'(\alpha)$ n'est pas nulle. Pour simplifier, supposons également que α n'est pas un vecteur mais un paramètre unique. Puisqu'il n'existe pas α pour lequel $\rho(\alpha_0)=0$, on fera des itérations. Pour cela, on se donne une valeur initiale α_0 et on cherche le plan tangent en α_0 à la fonction $\rho'(\alpha_0)$ ce qui est formellement équivalent à :

$$\rho(\alpha) \approx \rho(\alpha_0) + (\alpha - \alpha_0) \rho'(\alpha_0) \quad (\text{Cf. formule de Taylor à l'ordre 1})$$

En dérivant membre à membre, on obtient :

$$\rho'(\alpha) \approx \rho'(\alpha_0) + (\alpha - \alpha_0) \rho''(\alpha_0) \quad (\text{E 3.26})$$

Le terme de droite est une approximation de $\rho'(\alpha)$ et l'on peut approcher la racine de $\rho'(\alpha)$ par celle de $\rho'(\alpha_0) + (\hat{\alpha} - \alpha_0) \rho''(\alpha_0) = 0$

On a $\rho'(\alpha_0) + (\alpha - \alpha_0) \rho''(\alpha_0) = 0$, d'où :

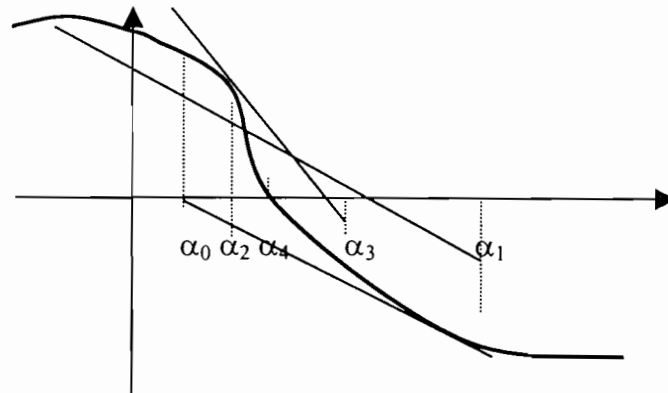
$$\alpha_1 = \alpha_0 - \frac{\rho'(\alpha_0)}{\rho''(\alpha_0)} \quad (\text{E 3.27a})$$

La démarche précédente est reprise en supposant α_1 comme valeur initiale cette fois-ci. Elle est substituée à α_0 dans (E 3.27a) et la démarche continue ainsi de suite. Les différentes itérations nous conduisent à la récurrence suivante :

$$\alpha_s = \alpha_{s-1} - \frac{\rho(\alpha_{s-1})}{\rho''(\alpha_{s-1})} \quad (\text{E 3.27b})$$

Lorsque la suite α_s converge vers une limite finie, celle-ci étant la solution du maximum de vraisemblance.

Graphique 2 : Illustration de la procédure itérative de Newton-Raphson.



Source : Andersen, page 47

Le graphique 2 illustre la méthode qui vient d'être décrite. En abscisse, se trouvent les valeurs de α_s obtenues à des itérations successives et en ordonnée les valeurs $\rho(\alpha_s)$ correspondantes. On remarque la solution initiale α_0 (point tangent à $\rho(\alpha)$ qui ne maximise pas encore $\rho(\alpha)$). A la première itération, on se situe en α_1 toujours tangent à $\rho(\alpha)$ et il y a forte perturbation de $\rho(\alpha)$ entre les deux solutions. Cette perturbation de $\rho(\alpha)$ s'atténue à partir de α_3 où il y a amorce d'un processus de convergence

On suppose maintenant que α au lieu d'être une constante, est un vecteur c'est-à-dire $\alpha=(\alpha(1), \alpha(2), \dots, \alpha(u))$. En itérant comme tout à l'heure, on arrive à la récurrence

$$\alpha_s = \alpha_{s-1} - Q(\alpha_{s-1})A^{-1}(\alpha_{s-1}) \quad (\text{E 3.28})$$

Où $Q(\alpha_s)$ est le vecteur gradient à u composantes dont les éléments sont les :

$$\rho'_j(\alpha_{s-1}) \quad (j = 1, \dots, u)$$

$$Q(\alpha_{s-1}) = (\rho'_1(\alpha_{s-1}), \rho'_1(\alpha_{s-1}), \dots, \rho'_1(\alpha_{s-1}))$$

$A(\alpha_{s-1})$ est la matrice hessienne (4x4) ; ses éléments sont les $a_{jr}(\alpha_{s-1})$ tels que :

$$a_{jr}(\alpha_{s-1}) = \rho''_{jr}(\alpha_{s-1})$$

De même si $\lim_{s \rightarrow \infty} \alpha_s = \hat{\alpha}$ alors $\hat{\alpha}$ est l'estimateur du maximum de vraisemblance

avec :

$$\hat{\alpha} = (\hat{\alpha}(1), \hat{\alpha}(2), \dots, \hat{\alpha}(u))$$

Quelques propriétés sont à énoncer ici :

- S'il existe un estimateur du maximum de vraisemblance $\hat{\alpha}$ alors cet estimateur est unique (Haberman 1974). Bandorf-Nielsen(1973) a montré plus particulièrement que les distributions appartenant à une famille exponentielle, à quelques exceptions près, avaient des estimateurs de maximum de vraisemblance uniques.

- Soient deux modèles M_1 et M_2 avec $M_1 \subset M_2$. Alors s'il existe un estimateur de maximum de vraisemblance pour M_2 , il en existe forcément pour M_1 . De même, s'il n'existe pas d'estimateur de maximum de vraisemblance pour M_1 , il n'en existera pas non plus pour M_2 . Cette propriété est surtout importante pour les modèles hiérarchique(Cf. chap. 4).

Nous ne pouvons clore ce chapitre sans toutefois nous pencher sur la présentation de quelques modèles linéaires. L'intérêt majeur de cette présentation est de dégager leur philosophie d'ensemble afin de mieux situer le modèle log-linéaire qui reste le guide de notre étude. Encore faudra-t-il noter que la prééminence que nous accordons au modèle log-linéaire n'est pas l'aboutissement d'un choix délibéré, mais s'inscrit dans une logique de traitement, une méthodologie dont la principale faculté est de mettre en relief les différents aspects structurels d'un tableau de contingence quelconque et qui, plus particulièrement dans le cadre des investigations à entreprendre ici, se prête mieux à l'analyse des tableaux à variables ordinales (Cf. chapitre 5).

3.5 PRESENTATION DE QUELOUES MODELES DE REGRESSION.

A l'origine, les premiers modèles linéaires ont été construits dans un but purement explicatif et prévisionnel. Ils présentent en général dans ce cas deux caractéristiques fondamentales : d'une part, les éléments déterministes et, d'autre part, les éléments aléatoires ou stochastiques. Ceux-ci peuvent varier et prendre un ensemble de valeurs si on se donne une population quelconque, alors que les premiers gardent leurs valeurs constantes.

Bien que initialement conçus pour expliquer et prévoir, les modèles peuvent être également utilisés pour la description des phénomènes. On parlera alors de modèles descriptifs. Ils sont définis "comme un processus de mise en correspondance d'un ensemble de données avec un ensemble de valeurs tirées d'un petit ensemble de paramètres, un critère pertinent étant utilisé pour déterminer la fidélité de la correspondance" (Nelder, Eurostat, 1984). Ces modèles ne font pas apparaître les relations de cause à effet entre les variables. Ils sont dits iconiques. C'est le cas bien entendu du modèle log-linéaire (Cf. chap. 4). Afin d'appréhender son approche et le situer par rapport aux modèles explicatifs, nous examinerons successivement dans les paragraphes 3.5.1, 3.5.2, 3.5.3 et 3.5.4 :

- les modèles de régression linéaire ;
- le modèle de régression multiplicative ;
- le modèle d'analyse de la variance ;
- la synthèse des modèles et situation du modèle log-linéaire.

3.5.1. Modèles de régression linéaire

Le modèle de régression linéaire est un modèle qui met en liaison une variable endogène (expliquée) avec une ou plusieurs variables exogènes (explicatives). Le modèle est ici linéaire des paramètres et non des observations. Nous présenterons le modèle de régression linéaire classique et le modèle de régression polynomiale.

3.5.1.1 Modèle de régression linéaire classique

Un modèle de régression linéaire classique sera caractérisé par une matrice à double entrée dans laquelle l'un des indices représente les observations et l'autre les variables explicatives. Ainsi, $t(i_1, i_2)$ représente la valeur prise par la $i_1^{\text{ème}}$ observation de la $i_2^{\text{ème}}$ variable explicative ($i_1 = 1, \dots, p$; $p \leq n$). On rappelle que $t(i_1, i_2)$ n'est pas de même nature que les observations d'une cellule élémentaire issues des tableaux multicroisés que nous avons vus (Cf. tableaux 1, 2, 3, 4).

Le modèle de régression linéaire classique est de la forme :

$$y(i) = \beta_0 + \beta_1 t(i, 1) + \beta_2 t(i, 2) + \dots + \beta_p t(i, p) + \varepsilon_i \quad (\text{E 3.29a})$$

Il y a donc un ensemble de n équations linéaires à $(p+1)$ inconnues (les p paramètres et le terme constant β_0).

(E 3.29a) permet d'expliquer la variable dépendante $y(i)$ grâce à une combinaison linéaire des variables explicatives. Matriciellement, (E3.29a) s'écrit :

$$Y = T\beta + \varepsilon$$

y est le vecteur $(n, 1)$ des valeurs expliquées ;

β est le vecteur $(p+1, 1)$ des paramètres à estimer ;

T est la matrice $(n, p+1)$ des observations prises par les variables exogènes ;

ε est le vecteur $(n, 1)$ des résidus du modèle.

En appliquant les hypothèses de Gauss-Markov et le principe d'estimation par la méthode des moindres carrés, on obtient : $\hat{\beta} = (T' T)^{-1} T' Y$.

Il s'agit d'un lissage, le principe consistant à rechercher une droite dans le plan (respectivement un hyperplan dans l'espace) qui représente le mieux le nuage de points du tableau.

3.5.1.2 Modèle de régression polynomiale

Le modèle de régression polynomiale prend la forme suivante :

$$Y(i_1) = \beta_0 + \beta_1 t(i_1, 1) + \beta_2 t^2(i_1, 2) + \dots + \beta_p t^p(i_1, p) + \varepsilon_{i_1} \quad (E 3.31)$$

$$i_1 = 1, \dots, n$$

On a toujours un modèle linéaire en β_j ($j = 0, \dots, p$) bien qu'apparaissent des monômes d'ordre 0 à p dans (E3.31). Le modèle est dit d'ordre p , et la procédure d'estimation est la même qu'en (E3.30). Cependant, les éléments de la matrice T ne sont pas de même nature que ceux de la matrice T de (E3.29) car ici, les observations sont plutôt élevées à des puissances, ce qui peut par conséquent provoquer l'altération de la qualité du modèle ((E 3.31) n'est pas toujours bon), à cause de l'existence possible des collinéarités entre les éléments de T . C'est pourquoi il serait préférable d'orthogonaliser ($T'T$) pour rendre indépendants les paramètres. Cette orthogonalisation donne un modèle de régression polynomiale orthogonale (le modèle est constitué de monômes orthogonaux) surtout utilisé pour des régressions où des variables explicatives prennent des valeurs à des intervalles de même longueur ou des données groupées en classes de même longueur.

Certaines observations peuvent être décrites par le produit de deux polynômes orthogonaux et si on en retient deux des degrés k et p ($i_1 = 0, \dots, k-1$; $i_2 = 0, \dots, p-1$) respectivement, on a :

$$y(i_1, i_2) = \sum_{s=0}^{k-1} \sum_{v=0}^{p-1} a_{s1} a_{v2} t^s(s) q^v(v)$$

$$s = 0, 1, 2, \dots, k-1 \quad (E3.32)$$

$$v = 0, 1, 2, \dots, p-1$$

L'observation $y(i_1, i_2)$ est simultanément expliquée par le produit de deux polynômes orthogonaux et on parle de double polynômes orthogonaux.

Le modèle (E 3.32) s'avère intéressant, ce d'autant plus qu'il pourra servir ultérieurement à la formulation du modèle log-linéaire appliqué aux variables ordinales (Cf. chap.5). Il reste que les polynômes orthogonaux comme les régressions polynomiales constituent toujours des méthodes de lissage qui font apparaître les évolutions moyennes des tendances et des courbes. D'autres formes de modèles seront projetées : les modèles de régression multiplicative

3.5.2 Modèle de régression multiplicative

Les formes linéaires des modèles précédents s'appliquent aux phénomènes dont les comportements sont linéaires. On retrouve pourtant des situations où le modèle linéaire est associé aux comportements multiplicatifs et c'est là que surviennent des incohérences. La forme idéale serait celle de l'association de chaque type de modèle aux phénomènes de même nature.

Le modèle de régression multiplicative que nous proposons ici a été construit par ZIGHERA et ce modèle tente de respecter l'aspect multiplicatif des phénomènes. Pour toute observation i_1 , il est tel que :

$$y(i_1) = k[t(i_1)]^a \quad (\text{E 3.33a})$$

En général, dans la statistique traditionnelle, on effectue une transformation conduisant à la linéarisation de (E 3.33a). Une telle transformation nous permet d'arriver au modèle :

$$\text{Log } y(i_1) = \text{Log } k + a \text{ Log } t(i_1) \quad (\text{E3.33b})$$

Lorsque la relation à établir correspond au modèle (E3.33a), le tableau des observations doit être tel que l'application de ce modèle conserve ses marges $t (+)$ et $y (+)$. D'autre part, les moyennes arithmétiques des $y (i_1)$ et des $t (i_1)$ ne devraient pas être affectées par le modèle. Malheureusement, quand intervient la transformation logarithmique (E3.33b), on obtient plutôt des moyennes géométriques qui n'ont pas de sens, de même que la sommation des

observations estimées ne permet pas d'aboutir aux différents totaux marginaux $t(+)$ et $y(+)$ du tableau initial. Afin de surmonter ces incohérences, on écarte le modèle (E3.33b) et on estime le modèle (E3.33a) sous les contraintes :

$$\sum y(i) = \sum \hat{y}(i); \quad \hat{y}(i) \text{ étant la valeur estimée de } y(i_1).$$

On cherche alors un vecteur d'observations \hat{y} qui soit le plus "proche" de y . Le problème posé est celui de la minimisation de l'information discriminante (Cf. chap. IV) sous contraintes, c'est-à-dire:

$$\text{Min } N(y/\hat{y}) \quad \text{s.c. } y(+) = \hat{y}(+) \quad (\text{E 3.34})$$

t étant le vecteur des valeurs explicatives

Puis que $\hat{y} = k t^\alpha$, on a : $N(y/\hat{y}) = N(y/k t^\alpha)$, d'où :

$$N\left(\frac{y}{\hat{y}}\right) = \sum \frac{y(i)}{y(+)} \text{Log} \frac{\frac{y(i)}{y(+)}}{\frac{t(i_1)^\alpha}{\sum t(i_1)^\alpha}} \quad (\text{E 3.35})$$

Le problème (E 3.34) revient à minimiser (E 3.35) avec les mêmes contraintes. En résolvant (E 3.35), on tire :

$$N\left(\frac{y}{\hat{y}}\right) = \sum \frac{y(i)}{y(+)} \text{Log} \frac{\sum t(i_1)^\alpha \text{Log} t(i_1)}{\sum t(i_1)^\alpha} \quad \text{qui est l'équation implicite de } \alpha, \text{ qui est une}$$

élasticité.

La présentation des modèles s'achèvera sur celui de l'analyse de la variance dont la décomposition est d'ailleurs assez proche, comme nous le verrons plus tard de celle du modèle Log-linéaire.

3.5.3 Modèle d'analyse de la variance.

Le modèle d'analyse de la variance a été étudié par Anderson (1954), Scheffé 1959), Lebart-Fenelon (1979), Clifford-Lunnebord (1983) etc.. Ce modèle permet d'expliquer une variable dépendante par un groupe de variables indépendantes, toutes de nature qualitative. Il peut encore se mettre sous la forme (E3.29b), où T n'a plus la même signification (cette matrice est la matrice de désignation ; Cf. 3.4.2). De même est le vecteur des "effets" (Cf. 4.1.1) liés aux variables.

En appliquant le modèle d'analyse de la variance au tableau 1, on obtient la décomposition

$$E[t(i_1, i_2)] = \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \alpha_{12}(i_1, i_2) + \varepsilon(i_1, i_2)$$

$E[t(i_1, i_2)]$ est l'espérance de la cellule élémentaire $t(i_1, i_2)$; $\varepsilon(i_1, i_2)$ est le résidu avec :

$$E[\varepsilon(i_1, i_2)] = 0$$

$$E[\varepsilon'(i_1, i_2) \varepsilon(i_1, i_2)] = \sigma^2.$$

Le modèle (E3.36) est un modèle aléatoire du fait de l'inclusion du terme aléatoire $\varepsilon(i_1, i_2)$ dans sa structure ; $\alpha_1(i_1)$ et $\alpha_2(i_2)$ sont les effets de 1^{er} ordre associés à la 1^{ère} variable et à la 2^{ème} variable respectivement.

$\alpha_{12}(i_1, i_2)$ représente le terme d'interaction entre les variables 1 et 2. Il existe une infinité de solutions $\alpha_1(i_1)$, $\alpha_2(i_2)$ et $\alpha_{12}(i_1, i_2)$ satisfaisant le modèle (E 3.36). Pour rendre ces solutions uniques, on ajoute à (E 3.36) les contraintes suivantes :

$$\begin{aligned} \sum \alpha_1(i_1) &= 0 \\ \sum \alpha_2(i_2) &= 0 \\ \sum_{i_1} \alpha_{12}(i_1, i_2) &= \sum_{i_2} \alpha_{12}(i_1, i_2) = 0 \end{aligned} \quad (\text{E 3.37})$$

Nous nous limitons à ce simple rappel du modèle qui suffit néanmoins à apporter d'autres éléments nouveaux par rapport aux modèles de régression prospectés plus haut. La synthèse d'ensemble de ces modèles pourra faciliter la situation du modèle Log-linéaire voire contribuer à un début de son élucidation.

3.5.4 Synthèse des modèles et situation du modèle Log-linéaire.

Les différents modèles que nous venons de voir sont des modèles linéaires, exception faite du modèle (E3.33) qui est multiplicatif et qui généralement se traite dans la statistique traditionnelle comme un modèle linéaire). Ces modèles permettent de "lisser" les données, c'est-à-dire approximer un nuage de points par une droite dans le plan ou un hyperplan dans l'espace (cas des modèles (E 3.29), (E3.36) ou par des courbes (Cf. (E3.31), (E3.32), (E3.33)).

Le modèle d'analyse de la variance, bien qu'explicatif, n'effectue pas ce genre de lissage. Les paramètres estimés dans ce modèle permettent d'évaluer la contribution des facteurs à la variable dépendante.

L'une des hypothèses retenues pour ces modèles est qu'ils suivent tous la loi normale. En sera-t-il autant pour d'autres modèles ? C'est là qu'apparaît l'une des différences avec le modèle Log-linéaire.

Le modèle Log-linéaire est avant tout un modèle linéaire dont la décomposition est identique à celle de l'analyse de la variance, avec deux grandes différences : il ne comporte pas de terme aléatoire (ce n est pas un modèle aléatoire) et de plus, c'est le logarithme de la valeur estimée d'une cellule d'un tableau quelconque qui se décompose en plusieurs effets. Bien que la décomposition des deux modèles soit identique, les différents effets dans les deux cas ne s'interprètent pas de la même façon. Le modèle Log-linéaire est descriptif et non explicatif comme les différents modèles que nous avons présentés plus haut. D'autre part, c'est un modèle multiplicatif en ce sens que dans un tableau, lorsque l'on veut comparer par exemple deux lignes quelconques, ce sont des rapports terme à terme des effectifs que l'on regarde et non des différences.

Le modèle Log-linéaire est également un modèle de "lissage". Ce lissage n'est pas à prendre de la même façon que celui défini pour les autres modèles linéaires que nous avons

présentés ici (Cf. 3.5). Il s'agit de réduire le nombre de cellules d'un tableau tout en conservant le maximum d'information de ce tableau. En d'autres termes, on projette un tableau ayant beaucoup de cellules et presque illisible sur un sous-espace qui le rend lisible. Le modèle Log-linéaire se résume à un modèle multiplicatif à contraintes additives puisque le lissage se fera sous la contrainte de la conservation des marges. Nous le verrons plus explicitement au chapitre IV.

CONCLUSION

Bien que non exhaustifs, les principaux concepts que nous venons de survoler dans le présent chapitre constituent pour le moins un cadre pouvant servir de base à la présentation du modèle Log-linéaire. Toutefois, nous aurons à définir, chaque fois qu'il sera possible, d'autres concepts d'analyse dans les chapitres ultérieurs. Tout en se gardant de hiérarchiser les différents outils proposés, il reste que les notions de multidimensionnalité des tableaux de contingence, de totaux marginaux, de marges, seront très couramment utilisées, les deux dernières servant même implicitement ou explicitement à la formulation du modèle (soit comme contraintes, soit comme marges associées à un effet), ce qui nous amène à poser le problème suivant : quelle est la forme structurelle du modèle Log-linéaire ? Sur quelle théorie est elle fondée au-delà du cadre d'analyse défini ? Applique-t-on indifféremment ce modèle aux différentes variables de nature qualitative ? Autant de questions auxquelles il faudrait apporter des réponses dont nous trouverons quelques éléments au chapitre IV.

CHAPITRE IV

MODELISATION LOG-LINEAIRE : FORMULATION DE BASE ET SON EXTENSION PAR LE BIAIS DU CENTRAGE DES PARAMETRES

4.0 INTRODUCTION

Nous avons défini (Cf. chap. 3) certains outils préliminaires susceptibles d'être appliqués aux modèles linéaires en général, et qui seront plus particulièrement utilisés pour l'étude du modèle Log-linéaire dans le présent chapitre. Dans ce modèle, la décomposition s'apparente à celle du modèle ANOVA, et le logarithme de l'observation à l'intérieur d'une cellule élémentaire d'un tableau de contingence p -dimensionnel ($p > 2$) est linéaire d'un ensemble de paramètres, d'où la dénomination Log-linéaire. Ce modèle est donc descriptif, et a pour principale caractéristique sa faculté à mettre en relief un ensemble de relations et d'interactions possibles entre les différentes variables qui constituent le tableau. Il est possible de prospecter ces relations au moyen des paramètres du modèle qui servent à formuler et tester les hypothèses. Ainsi, peuvent être testées les hypothèses d'indépendance totale, d'indépendance conditionnelle, de dépendance, etc., entre les différentes variables ou groupes de variables du tableau.

L'autre vocation du modèle Log-linéaire, c'est son adaptation au "lissage" d'un tableau. Dès lors, comment lisser -réduire le nombre de paramètres à estimer du modèle- le tableau de façon à adapter les paramètres restants à "l'information disponible du tableau ? Il apparaît ici la notion d'information et plus implicitement celle de la théorie de l'information qui sous-tend le modèle Log-linéaire. Nous évoquerons brièvement quelques aspects de son évolution depuis Fisher(1921) qui l'utilise pour la réduction des données (il parle de "l'information pertinente") jusqu'à nos jours en passant par l'entropie de Shannon, dont l'information discriminante (Cf. 4.1.2.), à la base du modèle Log-linéaire, est l'une des composantes de son extension.

Ce chapitre se présente en trois parties :

-tout d'abord, nous avons un premier volet dans lequel nous présentons le modèle Log-linéaire de base, notamment sa forme structurelle (4.1.3), celle-ci étant obtenue par le canal de l'information discriminante (principe de minimisation de l'information discriminante sous contraintes et plus particulièrement les contraintes internes (Cf. 3.4.2)) Les concepts d'effets caractérisant la structure du modèle et d'information dans un tableau voire de l'information discriminante auront été définis préalablement.

-ensuite, nous aborderons un second volet s'ouvrant sur la procédure d'estimation (4.2) et

reposant surtout sur le principe du maximum de vraisemblance. Ce principe comporte deux approches possibles, toutes les deux requérant l'existence d'une configuration minimale suffisante (4.2.1) l'approche directe et l'approche indirecte.

-l'estimation par l'approche indirecte peut être effectuée grâce à l'algorithme d'ajustement itératif (4.2.3.1) ou celui de Newton-Raphson (4.2.3.2). Toutefois, il existe d'autres méthodes d'estimation (4.2.3) que nous rappellerons brièvement : la méthode du score, la méthode des moindres carrés pondérés et celle du Khi-deux minimum.

-enfin, nous développons dans le troisième volet les techniques de choix des modèles Log-linéaires (4.3.1) et également l'extension de ces modèles (4.3.2). Les techniques de choix de modèles reposent sur de nombreux critères dont nous n'en retiendrons que deux.

Le premier critère est celui fondé sur les statistiques d'ajustement des modèles (4.3.1.1) et qui permettent initialement d'effectuer un choix entre deux modèles quelconques. Cependant, ces statistiques ont un certain nombre de limites, qui peuvent néanmoins être surmontées grâce aux indices d'ajustement (4.3.1.2). La procédure *stepwise* (4.3.1.3), dotée de ses deux démarches - la démarche "ascendante" et la démarche "descendante" est une généralisation du choix de modèles fondé sur ce premier critère.

Le second critère est celui du choix du modèle guidé par les termes standardisés des paramètres (4.3.1.4).

Dans le modèle Log-linéaire de base, les paramètres n'ont pas de signification importante, d'où l'intérêt à la base de son extension, qui reste tributaire de la technique de "centrage" des paramètres (4.3.2). C'est ainsi que les paramètres "centrés" permettront d'allouer (de partitionner) l'information totale d'un tableau entre les différentes marges indépendantes qui le constituent ; de même ils auront désormais un sens concret puisqu'ils serviront par exemple à positionner les modalités des variables.

4.1 PRESENTATION DU MODELE LOG-LINEAIRE A TRAVERS LA THEORIE DE L'INFORMATION

Ce chapitre sera dominé par le modèle Log-linéaire dont on explicitera surtout la forme structurelle de base avant d'étoffer les différentes procédures d'estimation et de choix de modèles qui lui sont rattachées.

La forme structurelle du modèle Log-linéaire est d'une importance capitale : elle permet de déceler, voire d'élucider les multiples structures latentes pouvant exister à l'intérieur d'un tableau multicroisé, ce par l'intermédiaire des termes "d'effets" servant à caractériser le modèle. Il est donc judicieux, avant d'explorer le modèle proprement dit, de préciser davantage la notion d'effet, ce que nous essayerons de faire à travers quelques exemples concrets que l'on peut rencontrer dans un tableau de contingence.

Dans ce paragraphe, nous tenterons d'examiner assez brièvement le concept d'effet ; ensuite nous rappellerons la notion de l'information dans un tableau et par là même ses différentes variantes (entropie, négentropie, information discriminante), cette notion sous-tendant le modèle Log-linéaire que nous présenterons pour terminer en montrant comment traditionnellement, on détermine les différents effets.

4.1.1 Notion et signification des effets.

Si l'on se donne par exemple un tableau qui croise les effectifs féminins suivant les régions et les métiers (il s'agit bien entendu des données françaises), on peut constater l'existence d'une forte participation féminine dans les régions dominées par les métiers du tertiaire et une faible participation féminine dans les régions dominées par les métiers du secondaire, ce d'autant plus que l'emploi féminin global est surtout concentré dans le tertiaire. Il y a donc un effet de structure qui se caractérise par cette différence dans la participation féminine. De même, si l'on mesure globalement le taux d'activité des jeunes de 25 à 30 ans pour toutes les régions données, et que l'on constate qu'il est assez élevé, alors les régions où cette tranche d'âge est largement représentée auront par rapport à celles où elle est peu représentée des taux d'activité assez grands pour la tranche d'âge considérée. Il y a encore ici une différence entre les taux d'activité régionaux qui reste tributaire de la répartition régionale par tranches d'âge. Ces deux exemples peuvent aider à mieux comprendre la notion d'effet caractérisant la structure du modèle, voire du tableau. Plusieurs types d'effets se distinguent dans un modèle et chaque type sera lié à une marge du tableau : effet de 1^{er} ordre, effet de 2nd ordre, ..., effet d'ordre p (p est la dimension du tableau).

-Effet du premier ordre.

L'effet de 1^{er} ordre est l'effet associé à chaque marge unidimensionnelle (marge d'ordre

1) du tableau p-dimensionnel. Cet effet marque la différence entre les observations du tableau, et reflète la différence existant entre les marges d'ordre 1 de ce tableau. Ainsi, si l'une des variables du tableau est le sexe, les observations liées à la modalité "femme" seront plus importantes que celle liées à la modalité "homme" s'il y a moins d'hommes que de femmes dans l'échantillon permettant de ventiler ce tableau. Si la variable sexe est la variable V_1 alors cet effet s'écrit : $\alpha_1(i_1)$.

- Effet du deuxième ordre.

En présence de deux ou davantage de variables, on peut calculer l'effet de second ordre qui traduit la différence à l'intérieur des observations, cette différence étant le reflet cette fois-ci de la corrélation entre les deux variables ou toutes les paires de variables. L'effet de 2nd ordre entre les variables 1 et 2 du tableau est noté : $\alpha_{1,2}(i_1, i_2)$.

Ainsi, soient deux variables d'un tableau telles que le "niveau d'étude" et le "niveau de salaire". Alors les observations à l'intersection des modalités "haut niveau d'étude" et "haut niveau de salaire" seront relativement plus importantes que les autres s'il y a corrélation entre "niveau d'étude" et "niveau de salaire".

- Effet du troisième ordre.

L'effet du 3^{ème} ordre mesure pour trois variables croisées dans un tableau, la différence entre les observations de ce tableau et qui reflète une interaction (spécification).

Si l'on croise les deux variables précédentes suivant les "régions" et qu'il existe une différence importante dans "l'île de France" (spécification) pour les modalités "haut niveau d'étude" et "haut niveau de salaire" (hypothèse retenue d'une corrélation entre "niveau d'étude" et "niveau de salaire"), alors l'effet du 3^{ème} ordre se traduira par l'importance plus grande des observations à l'intersection des modalités "haut niveau d'étude", "haut niveau de salaire" et "île de France".

Cette mesure d'effet se généralise facilement à l'ordre supérieur et chaque fois l'effet d'ordre p traduira une interaction avec des interprétations de plus en plus lourdes. L'effet du 4^{ème} ordre par exemple marque la différence entre les observations et qui reflète une interaction (spécification) d'une interaction (effet d'ordre 3 bien évidemment). Exemple : la différence du niveau du salaire suivant le niveau d'étude dans l'île de France dépend de la profession.

Après avoir essayé d'éclairer au mieux la notion d'effet qui apparaîtra dans les modèles Log-linéaire, on pourra se pencher sur celle de l'information -de la théorie de l'information- qui reste l'un des fondements de la modélisation Log-linéaire.

4.1.2 Notion d'information dans un tableau.

Tout tableau statistique procure une information à celui qui l'exploite et la question qu'on peut se poser est de savoir si cette information est quantifiable. C'est ce problème que nous aborderons ici en examinant la mesure de la quantité d'information dans un tableau et celle de l'information discriminante, non sans avoir précisé au départ quelques éléments de la théorie de l'information.

4.1.2.1 Quelques éléments de la théorie de l'information.

La dénomination "théorie de l'information" fait penser de prime abord à l'existence d'une théorie propre à l'information. Toute théorie étant fondée sur une axiomatique, "la théorie de l'information", en ce sens supposerait la définition d'un ensemble d'axiomes lui servant de support. Cependant, cette dénomination est équivoque, et comme le souligne P. J. Lancry, «la théorie de l'information n'est pas une théorie de l'information [...] (et) recouvre des champs divers dont les concepts de base sont aussi abondants et polymorphes que l'imagination des chercheurs».

Ce polymorphisme illustre de même qu'il reflète une certaine évolution qu'a connue la théorie de l'information depuis Fisher et surtout Shannon jusqu'à nos jours. En effet, Fisher (1921) fut le premier à développer formellement et théoriquement la notion de l'information dans le but de déterminer statistiquement les paramètres d'une loi de probabilité permettant de "mieux" représenter des données. Il y a chez Fisher un souci dominant qui est celui de la recherche de la réduction des données, ce qui l'amène à parler de "l'information pertinente" qu'on pourrait définir comme la part utile de l'information extraite de cette réduction.

Le développement de la théorie de l'information prendra de l'ampleur dès 1948 à la suite des travaux de Shannon. Sa théorie est avant tout une théorie générale de la communication. Il définit une quantité d'information $\sum p_i \log p_i$ (p_i est la probabilité associée

à un état d'un système) à laquelle il donne le nom d'entropie. Les applications qui découlent de cette notion d'entropie sont nombreuses, notamment avec Jaynes(1957) et Evans(1962) qui l'utilisent en thermodynamique statistique (certes, les deux auteurs ne sont pas les premiers à l'utiliser en thermodynamique puisque les premières applications de l'entropie à la thermodynamique sont encore plus anciennes). Reilly et Blau(1974) qui l'utilisent en chimie, P. Thionnet (1963) et surtout Theil(1967) dont le champ d'application s'étend à l'économie. Theil lui donnera le nom de "contenu informatif". Il reste que l'entropie est une grandeur absolue, alors que "l'information, à travers ses aspects objectifs et subjectifs, c'est-à-dire en tant qu'élément de base d'une théorie de la forme et de la signification est avant tout. une notion relative Cette notion de relativité est traduite par une situation où l'on a deux alternatives l'une a priori et l'autre a posteriori. Ainsi, en présence de deux alternatives, l'information que l'on tire de la connaissance a posteriori est $\sum p_i \text{Log}(p_i/q_i)$ (p_i est la probabilité de l'évènement a priori et q_i celle de l'évènement a posteriori). On l'appelle information Kullbackienne ou information discriminante. Initialement définie par Wiener(1948), celle-ci a été largement reprise par Kullback(1959, 1978), Johnson(1979), Shore(1980,1981), pour ne citer que ceux-là, dans le cadre des applications statistiques. On lui prête d'autres appellations, notamment, divergence informationnelle, entropie relative, entropie croisée, déviance. Nous retiendrons pour la suite la dénomination information discriminante. Avant d'introduire le modèle Log-linéaire qui est sous-jacent a l'information discriminante, il est nécessaire de faire un détour pour voir comment s'apprécie la quantité d'information dans un tableau.

4.1.2.2. Quantité d'information dans un tableau quelconque.

Soit T un tableau de contingence p-dimensionnel avec $t(i_\omega)$ comme observation à l'intérieur d'une cellule(i_ω) de ce tableau ((i_ω) représente p-uplets d'indices soit unidimensionnels, soit multidimensionnels). On suppose que le tableau t a I cellules et que la contrainte interne satisfaisant la marge d'ordre zéro est vérifiée c'est-à-dire : $\sum t(i_\omega) = t(+)$

La mesure de la quantité d'information du tableau, notée $N(T)$ est définie par :

$$N(T) = \sum \frac{t(i_\omega)}{t(+)} \text{Log} \frac{t(i_\omega)}{t(+)} \quad (\text{E4.1})$$

Cette quantité d'information s'appelle négentropie ou encore indicateur de Theil(1967), celui-ci l'utilisant comme indicateur de la mesure de l'inégalité des revenus. La négentropie sert alors à mesurer la dispersion d'un tableau et par conséquent, peut être considérée comme un coefficient de variation.

Lorsque le tableau d'étude est plat, c'est-à-dire qu'il est caractérisé par l'équirépartition (toutes les valeurs $t(i_{\omega})/ t(i_{\omega})$ sont identiques quelle que soit la cellule (i_{ω})), alors est atteinte la borne inférieure de $N(T)$. A l'inverse si toutes les observations du tableau sont concentrées dans -une seule cellule, les cellules restantes étant vides, $N(T)$ prend sa plus grande valeur $\text{Log} I$ de telle sorte que l'inégalité suivante est vérifiée .

$$0 < N(T) < \text{Log} I \quad (\text{E } 4.2)$$

Par conséquent, la négentropie sera d'autant plus grande que la taille I du tableau est également grande. En d'autres termes, plus le tableau d'étude sera très grand, plus il sera riche en information. Mais cette information risque d'être inaccessible et difficile à interpréter du fait de la complexité de la structure du tableau. Il est alors souhaitable de le projeter dans un espace lisible, où l'on peut acquérir le maximum d'information. Cette projection revient à rechercher un tableau T' qui ressemble le plus à T (nous y reviendrons plus loin), c'est-à-dire qui est structurellement "proche" de T . Cette proximité se mesure d'ailleurs au moyen de l'information discriminante.

4.1.2.3 L'information discriminante entre deux tableaux.

On peut étendre la notion d'information définie en (E 4.1), laquelle sera appréhendée comme une mesure d'écart entre deux tableaux (deux distributions) quelconques T et T' . Cet écart est l'information discriminante entre T et T' , et notée $N(T/T')$.

En considérant $t(i_{\omega})$ et $t'(i_{\omega})$ comme les observations à l'intérieur des cellules de T et T' (on suppose que T' a le même nombre de cellules que T) respectivement avec :

$$\sum t(i_{\omega}) = \sum t'(i_{\omega}) = t(+)$$

$$N(T/T) = \sum \frac{t(i_\omega)}{t(+)} \text{Log} \left[\frac{\frac{t(i_\omega)}{t(+)} / \frac{t'(i_\omega)}{t(+)}}{\frac{t(i_\omega)}{t(+)} / \frac{t'(i_\omega)}{t(+)}} \right]$$

$$= \sum \frac{t(i_\omega)}{t(+)} \text{Log} \frac{t(i_\omega)}{t'(i_\omega)} \quad (\text{E4.3})$$

On constate alors que l'information discriminante est une moyenne pondérée des logarithmes des rapports $t(i_\omega)/t'(i_\omega)$.

Lorsque T et T' ont même "structure" (mêmes observations dans chaque cellule), on obtient la valeur minimale de $N(T/T')$ qui est zéro et $N(T/T')$ devient infinie si au moins une des cellules du tableau T' est nulle, ce qui nous donne le domaine de variation de $N(T/T')$ suivant:

$$0 \leq N(T/T') \leq \infty \quad (\text{E4.4})$$

Remarque : L'écart défini en (E 4.3) est de nature multiplicative et cet écart n'est pas une distance, car au moins une propriété de la distance, en l'occurrence celle de la symétrie n'est pas vérifiée. En effet $N(T/T') \neq N(T'/T)$. Il existe néanmoins une information discriminante satisfaisant la propriété de la symétrie : c'est le cas de l'information discriminante symétrisée ou divergence.

D'autre part, $N(T/T')$ telle qu'exprimée en (E4.3) est définie par le biais d'une distribution considérée être celle des probabilités, puisqu'on a :

$$\sum \frac{t(i_\omega)}{t(+)} = 1, \text{ de même que } \frac{t(i_\omega)}{t(+)} = 0$$

$$\text{Donc, } 0 \leq \frac{t(i_\omega)}{t(+)} \leq 1$$

(E 2.3) pourra s'exprimer également en termes d'effectifs, ce qui se justifie par la formulation suivante:

$$N(T/T') = \sum t(i_{\omega}) \text{Log} \frac{t(i_{\omega})}{t(i_{\omega})} \quad (\text{E 4.3a})$$

$N(T/T')$ suit approximativement un Khi-deux. On peut l'utiliser pour comparer la structure totale d'un tableau avec une structure hypothétique. Il se profile alors l'idée que nous avons avancée plus haut et qui étayait la possibilité de rechercher un tableau T' "proche" structurellement de T . De cette idée, apparaît également la philosophie inhérente au modèle Log-linéaire. Cette similitude n'est pas d'ailleurs neutre et nous verrons dans ce qui suit, que la recherche d'une structure hypothétique en s'appuyant sur l'information discriminante nous conduit naturellement au modèle log-linéaire.

4.1.3 Approche introductive et forme structurelle du modèle Log-linéaire.

Nous examinerons ici le modèle Log-linéaire (cas du modèle saturé et du modèle non saturé) et la procédure traditionnelle de calcul des effets de paramètres, lesquels caractérisent la forme structurelle du modèle. Auparavant, nous aurons ébauché le principe d'introduction du modèle Log-linéaire.

4.1.3.1 Exploration de la logique d'ensemble et introduction de la modélisation Log-linéaire.

On désigne par T un tableau de contingence quelconque. L'un des problèmes auxquels se trouvent confrontés les statisticiens est celui de l'existence possible d'un volume important de données, largement réparties sur un grand nombre de cellules de ce tableau. Ce foisonnement de données aura pour conséquence de rendre difficilement accessible l'information totale du tableau (on a vu avec le concept de négentropie que cette information était d'autant plus riche que la taille des cellules augmentait) même si l'on dispose de puissants moyens informatiques de nos jours. Outre le fait que l'accession à l'information totale du tableau est possible au gré d'une mobilisation d'importants moyens financiers qu'exige l'outil informatique, se pose surtout le problème de son interprétation. Est-il aisé d'interpréter une information à structure complexe?. Comment procéder pour la rendre facilement compréhensible, interprétable "lisible" ?

Il est donc nécessaire d'effectuer un déménagement de l'information totale du tableau de manière à en extraire la partie "utile" (concept d'information "pertinente" chez Fisher) et la restituer dans une structure plus simple. Ce déménagement s'accommode parfaitement avec la procédure de lissage d'un tableau et qui consiste à réduire le nombre de cellules du tableau (on perd de l'information) tout en conservant le maximum d'information de ce tableau. En d'autres termes, "on projette une information trop riche sur un espace lisible tout en minimisant la perte d'information" qui en découle. Cette projection équivaut à rechercher un tableau T' à I' cellules ($I' < I$) très "proche" de T , c'est-à-dire que l'écart (Cf. E4.3.a) entre T et T' doit être la plus petite possible. Formellement, nous minimisons cet écart.

D'autre part, la projection engendre une infinité de tableaux T' comme sous-partition de T . On retiendra parmi ces tableaux ceux qui respectent les contraintes internes (Cf. 3.4.2) c'est-à-dire les tableaux les plus "proches" de T . D'ailleurs, plus les contraintes internes seront nombreuses, plus on se rapprochera de T . Ces différents tableaux sont des tableaux simplifiés de T . On verra plus loin qu'ils permettent d'ajuster T .

Appelons dorénavant T^* le tableau simplifié. Ce tableau se situe donc entre les deux cas extrêmes que sont le tableau T et le tableau uniforme U qui a pour observation à l'intérieur d'une cellule élémentaire $x(i_\omega)$. Pour que soient satisfaites les contraintes internes, il faut que soit vérifiée la condition suivante :

$$\mathbf{t} \mathbf{t}' = \mathbf{t} \mathbf{t}^* \quad \text{E(4.5a)}$$

où \mathbf{t} et \mathbf{t}' représentent, respectivement le vecteur des observations -elles sont prises dans un ordre lexicographique- des tableaux T et T^* . On peut encore écrire

$$\sum_{i_\omega} \theta_\omega(i_\omega) \mathbf{t}(i_\omega) = \sum_{i_\omega} \theta_\omega(i_\omega) \mathbf{t}^*(i_\omega) \quad \text{(E4.5b)}$$

où $\theta_\omega(i_\omega)$ est l'ensemble des éléments de \mathbf{t} ($\theta_\omega(i_\omega) = 0$ ou 1); (i_ω) étant la cellule élémentaire associée à la marge pour la marge $M(i_1)$, $i_\omega = (i_1)$ et on peut écrire $\theta_1(i_1)$; pour la marge $M(i_1, i_2)$, $i_\omega = (i_1, i_2)$ et on écrit $\theta_{12}(i_1, i_2)$. Si le tableau est p -dimensionnel, il existe 2^p marges possibles avec $\theta_0(i_\omega) = 1, \forall i_\omega$.

Traditionnellement, pour évaluer T^* , on part de la situation la plus simple caractérisée par un tableau uniforme U et auquel on ajoute une ou plusieurs contraintes pour arriver à T^* "proche" de T . Compte tenu de tout ce qui a été dit précédemment, cela entraîne la résolution du problème suivant:

$$\begin{aligned} \text{Min } & \sum_{i_\omega} t^*(i_\omega) \text{Log} \frac{t^*(i_\omega)}{x(i_\omega)} \\ \text{s.c } & \sum_{i_\omega} \theta_\omega(i_\omega) y(i_\omega) = \sum_{i_\omega} \theta_\omega(i_\omega) y^*(i_\omega) \end{aligned} \quad (\text{E 4.6})$$

$$\omega = 1, 2, \dots, s$$

s est l'ensemble de toutes les marges possibles, C'est en résolvant (E 4.6), qu'on retrouve la forme structurelle du modèle Log-linéaire.

4.1.3.2. Forme structurelle du modèle Log-linéaire.

Soit α le vecteur des multiplicateurs de Lagrange ($\alpha_1(i_1), \alpha_2(i_2), \dots, \alpha_\omega(i_\omega)$) représente les composantes de α . Le Lagrangien L du problème (E4.6) est tel que :

$$L = \sum_{i_\omega} t^*(i_\omega) \text{Log} [t^*(i_\omega) / x(i_\omega)] + \sum_{i_\omega} \alpha_\omega(i_\omega) \left[\sum_{i_\omega} \theta_\omega(i_\omega) y(i_\omega) - \sum_{i_\omega} \theta_\omega(i_\omega) y^*(i_\omega) \right] \quad (\text{E 4.7})$$

La différenciation de (E 4.7) et la résolution des équations relatives aux conditions de 1^{er} ordre (condition d'existence d'un extremum) entraînent :

$$\text{Log} [t^*(i_\omega) / x(i_\omega)] = \sum_{i_\omega} \alpha_\omega(i_\omega) \theta_\omega(i_\omega) \quad (\text{E4.8})$$

$$\text{Log} t^*(i_\omega) = \text{Log} x(i_\omega) + \sum_{i_\omega} \alpha_\omega(i_\omega) \theta_\omega(i_\omega) \quad (\text{E4.9a})$$

Cette dernière équation donne le modèle Log-linéaire et indique que le logarithme naturel de chaque observation du tableau simplifié (par la suite nous l'appellerons tableau estimé) T^* est linéaire d'un ensemble de paramètres $\alpha_\omega(i_\omega)$ qui sont les multiplicateurs de Lagrange.

Posons $\text{Log } x(i_\omega) = \alpha_0$

$$\text{Log } t^*(i_\omega) = \alpha_0 + \sum_{\omega=1} \alpha_\omega(i_\omega) \theta_\omega(i_\omega) \quad (\text{E 4.9b})$$

Cette décomposition générale ne permet pas de distinguer le modèle saturé du modèle non saturé, alors qu'une telle distinction conforte davantage l'idée et l'objet de la projection évoquée plus haut.

4.1.3.2.1. Modèles saturés.

Un modèle saturé est un modèle qui, appliqué à un tableau de contingence a autant de paramètres indépendantes que le nombre totale de cellules de ce tableau. Ainsi, dans un tableau de dimension 2, le modèle saturé a la forme :

$$\text{Log } t^*(i_1, i_2) = \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \alpha_{12}(i_1, i_2) \quad (\text{E4.10})$$

(E4.10) signifie également que toutes les cellules du tableau sont prises en compte dans les contraintes du modèle. Cette décomposition est analogue à celle du modèle ANOVA (Cf. 3.5.3). On a un système d'équations en (E 4.10) et il est indéterminé (plusieurs paramètres satisfont (E 4.10)). Afin qu'il soit donc déterminé, on impose les contraintes linéaires suivantes aux effets de paramètres :

$$\begin{aligned} \sum_{i_1} \alpha_2(i_1) &= \sum_{i_2} \alpha_2(i_2) = 0 \\ \sum_{i_1} \alpha_{12}(i_1, i_2) &= \sum_{i_2} \alpha_{12}(i_1, i_2) = 0 \end{aligned} \quad (\text{E4.10})$$

Les contraintes (E4.11) caractérisent l'unicité des paramètres $\alpha_\omega(i_\omega)$ dans le sous-espace de projection. Dans un tableau de dimension 3 où l'on a croisé trois variables V_1, V_2 et V_3 ayant respectivement I_1, I_2 et I_3 modalités, le modèle saturé a la décomposition ci-après :

$$\begin{aligned} \text{Log } t^*(i_1, i_2, i_3) &= \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \alpha_3(i_3) + \alpha_{12}(i_1, i_2) + \alpha_{13}(i_1, i_3) + \alpha_{23}(i_2, i_3) + \alpha_{123}(i_1, i_2, i_3) \\ \forall i_1 &= 1, \dots, I_1; \quad \forall i_2 = 1, \dots, I_2; \quad \forall i_3 = 1, \dots, I_3. \end{aligned}$$

Le modèle saturé n'est donc qu'une reparamétrisation totale du tableau, c'est-à-dire qu'il permet la reconstitution intégrale du tableau. En sera-t-il autant des modèles non saturés ?

4.1.3.2.2. Modèles non saturés.

Un modèle non saturé est un modèle qui appliqué à un tableau possède moins de degré de contraintes que le nombre de cellules indépendantes du tableau. On retrouve encore l'idée forte qui guide la méthodologie Log-linéaire et développée en (4.1.3.1) en terme d'information : l'exploration des structures complexes du tableau par sa projection sur un espace "lisible". Ainsi, un modèle non saturé se déduit du modèle saturé par suppression d'un ou plusieurs termes d'effets de paramètres.

Pour un tableau de dimension 2, le modèle suivant est non saturé :

$$\text{Logt}^*(i_1, i_2) = \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) \quad (\text{E 4.14})$$

En effet, il a $I_1 + I_2 - 1$ degrés de contraintes et on constate qu'il se déduit de (E 4.10) (ce modèle a $I_1 \times I_2$ degrés de contraintes) par suppression du terme $\alpha_{12}(i_1, i_2)$. La nullité du terme $\alpha_{12}(i_1, i_2)$ traduit l'indépendance totale des variables du tableau.

Dans un tableau p-dimensionnel, un modèle non saturé permettra de formuler et tester les hypothèses d'indépendance ou de dépendance conditionnelle entre des variables du tableau. Du modèle (E 4.12), on a par exemple le modèle d'indépendance conditionnelle entre les variables V_2 et V_3 , connaissant la variable V_1 et qui se traduit par :

$$\text{Logt}^*(i_1, i_2, i_3) = \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \alpha_3(i_3) + \alpha_{12}(i_1, i_2) + \alpha_{13}(i_1, i_3) \quad (\text{E 4.15})$$

L'objet fondamental du modèle Log-linéaire sera donc de chercher un modèle non saturé qui ajuste au mieux les observations du tableau -des données brutes (Cf. 4.3). Nous n'avons prospecté qu'une classe particulière de modèles à savoir les modèles hiérarchiques. La caractéristique principale de ces modèles est qu'il ne peut y avoir d'effets d'ordre supérieur que s'il y a des effets d'ordre inférieur et vice-versa. C'est d'ailleurs sur la base de ces modèles

que l'on fondera les technique!; de choix de modèles Cette structure hiérarchique apparaîtra dans la procédure traditionnelle de calcul des effets de paramètres.

4.1.4 Procédure traditionnelle de calcul des paramètres.

Le calcul des effets de paramètres se fait de façon récursive puisque la détermination des paramètres d'ordre inférieur est indispensable pour le calcul des paramètres d'ordre supérieur.

Reprenons le modèle Log-linéaire dans sa forme générale telle que définie en (E 4.9b). Ce modèle a 2^p termes d'effets à estimer. Il est bien entendu qu'en cas de modèle non saturé, on aura moins de termes. De (E4.9b) le calcul des différents termes du modèle se fait suivant le principe ci-après :

$$\alpha_0 = \frac{1}{\prod_j} \sum \dots \sum \text{Log}t(i_1, i_2, \dots, i_p) = \text{Log}[\prod \dots \Pi t(i_1, \dots, i_p)]^{\frac{1}{\prod_j}} \quad (\text{E4.16})$$

$$\alpha_1(i_1) = \frac{1}{\prod_j} \sum \dots \sum \text{Log}t(i_1, i_2, \dots, i_p) - \alpha_0$$

$$= \text{Log}[\prod \dots \Pi t(i_1, \dots, i_p)]^{\frac{1}{\prod_j}} - \text{Log}[\prod \dots \Pi t(i_1, \dots, i_p)]^{\frac{1}{\prod_j}}$$

$$\alpha_k(i_k) = \text{Log}[\prod \dots \Pi t(i_1, \dots, i_p)]_{j \neq i}^{\frac{1}{\prod_j}} - \text{Log}[\prod \dots \Pi t(i_1, \dots, i_p)]^{\frac{1}{\prod_j}}$$

$$k = 1, 2, \dots, p \quad (\text{E 4.17})$$

$\alpha_k(i_k)$ étant l'effet de 1^{er} ordre associé à la variable V_k , on constate qu'il mesure un écart (déviations) par rapport à, l'effet moyen.

De même, $\alpha_{kl}(i_k, i_l)$ est tel que.

$$\begin{aligned}
\alpha_{k1}(i_k, i_1) &= \frac{1}{\prod_{j \neq k} I_j} \prod_{j \neq i} \dots \prod \text{Log} t(i_1, \dots, i_p) - \alpha_k(i_k) - \alpha_1(i_1) - \alpha_0 \\
&= \text{Log} \left[\prod_{j \neq i} \dots \prod t(i_1, \dots, i_p) \right]_{j \neq i}^{\frac{1}{\prod I_j}} - \text{Log} \left[\prod_{j \neq i} \dots \prod t(i_1, \dots, i_p) \right]_{j \neq i}^{\frac{1}{\prod I_j}} \\
&\quad - \text{Log} \left[\prod_{j \neq i} \dots \prod t(i_1, \dots, i_p) \right]_{j \neq i}^{\frac{1}{\prod I_j}} - \text{Log} \left[\prod_{j \neq i} \dots \prod t(i_1, \dots, i_p) \right]_{j \neq i}^{\frac{1}{\prod I_j}} \quad (\text{E 4.18}) \\
l &= 1, 2, \dots, p \\
k &= 1, 2, \dots, p
\end{aligned}$$

En général, soit ω une marge quelconque ($\omega \in \Omega$) et ω' toutes les marges impliquées de ω , alors $\alpha_{\omega}(i_{\omega})$ est tel que :

$$\alpha_{\omega}(i_{\omega}) = \text{Log} \left[\prod_{j \neq i} \dots \prod t(i_1, \dots, i_p) \right]_{j \neq i}^{\frac{1}{\prod I_j}} - \sum \alpha_{\omega'}(i_{\omega'}) - \alpha_0 \quad (\text{E 4.19})$$

$\alpha_{\omega}(i_{\omega})$ représente les effets associés aux différentes marges impliquées de ω .

Cette procédure de calcul des effets de paramètres se fait ici de manière plus directe alors qu'en général elle est liée aux observations lissées d'un tableau. Somme toute, les différents modèles Log-linéaires que nous venons d'examiner -plus particulièrement, les modèles non saturés- sont obtenus une fois qu'ont été estimées les observations du tableau (il existe néanmoins une démarche inverse). Il est alors fort intéressant d'explorer les méthodes d'estimations couramment utilisées.

4.2 TECHNIQUES D'ESTIMATION POUR UN MODELE DONNE.

Après avoir présenté le modèle Log-linéaire, nous sommes dorénavant au cœur du problème de son estimation, du moins dans le paragraphe qui s'ouvre ici. Cette estimation repose très généralement sur le principe du maximum de vraisemblance que nous avons

globalement abordé en 3.4 (Cf. chapitre 3). Les différents aspects de ce principe seront donc repris en vue de leur application particulière au modèle Log-linéaire.

La recherche des statistiques minimales suffisantes (4.2.1) sera d'abord effectuée avant d'entamer la procédure d'estimation proprement dite (4.2.2). Cette technique n'étant pas la seule, nous indiquerons au-delà d'autres méthodes d'estimation 4.2.3.

4.2.1 Recherche d'une configuration minimale suffisante.

Le modèle Log-linéaire est un modèle exponentiel (Cf. E4.8). On a également montré qu'un modèle exponentiel pouvait s'écrire comme fonction des statistiques suffisantes (Cf. E3.19). L'application du processus d'échantillonnage poissonnien (3.3.1) à un tableau de dimension 2 permet de déterminer une fonction de vraisemblance dont le logarithme est:

$$\sum_{i_1} \sum_{i_2} t(i_1, i_2) \text{Log} t^*(i_1, i_2) - \sum_{i_1} \sum_{i_2} t^*(i_1, i_2) - \sum_{i_1} \sum_{i_2} t(i_1, i_2) \quad (\text{E4.20})$$

De cette équation, seul le terme $\sum_{i_1} \sum_{i_2} t(i_1, i_2) \text{Log} t^*(i_1, i_2)$ dépend des paramètres, les deux termes restants n'étant que des constantes ($\sum_{i_1} \sum_{i_2} t(i_1, i_2) = N$). Donc maximiser (E 4.20) revient à maximiser ce terme. Si le tableau est lissé par le modèle saturé (E4.10), on arrive à

$$\begin{aligned} \sum_{i_1} \sum_{i_2} t(i_1, i_2) \text{Log} t^*(i_1, i_2) &= \sum_{i_1} \sum_{i_2} t^*(i_1, i_2) [\alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \alpha_{12}(i_1, i_2)] \\ &= \sum_{i_1} \sum_{i_2} \alpha_0 t(i_1, i_2) + \sum_{i_1} \sum_{i_2} \alpha_1(i_1) t(i_1, i_2) + \sum_{i_1} \sum_{i_2} \alpha_2(i_2) t(i_1, i_2) + \sum_{i_1} \sum_{i_2} \alpha_{12}(i_1, i_2) t(i_1, i_2) \quad (\text{E 4.21}). \end{aligned}$$

Après simplification, on obtient

$$\sum_{i_1} \sum_{i_2} t(i_1, i_2) \text{Log} t^*(i_1, i_2) = N\alpha_0 + \sum_{i_1} \alpha_1(i_1) t(i_1, +) + \sum_{i_2} \alpha_2(i_2) t(+, i_2) + \sum_{i_1} \sum_{i_2} \alpha_{12}(i_1, i_2) t(i_1, i_2) \quad (\text{E 4.22})$$

On conclue que les marges $M(0)$, $M(i_1)$, $M(i_2)$ et $M(i_1, i_2)$ sont des statistiques minimales suffisantes pour l'estimation des paramètres du modèle saturé. Le principe de la recherche de

ces statistiques est identique pour tous les modèles saturés lissant un tableau p-dimensionnel. Cependant, la recherche de la statistique minimale suffisante trouve son intérêt majeur lorsque l'on est en présence d'un modèle non saturé. On peut donc déduire de (E 4.12) un modèle non saturé par suppression des effets $\alpha_{23}(i_2, i_3)$ et $\alpha_{23}(i_2, i_3)$ et tel que :

$$\text{Log } t^*(i_1, i_2, i_3) = \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \alpha_3(i_3) + \alpha_{12}(i_1, i_2) + \alpha_{13}(i_1, i_3) \quad (\text{E4.23a})$$

qui est encore équivalent à :

$$\text{Log } t^*(i_1, i_2, i_3) = \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \alpha_3(i_3) + \alpha_{12}(i_1, i_2) + \alpha_{13}(i_1, i_3) - \alpha_1(i_1) + \alpha_1(i_1)$$

$$\text{Log } t^*(i_1, i_2, i_3) = \alpha_0 + C_{12} + C_{13} - R_1 \quad (\text{E 4.23b})$$

C_{12} et C_{13} sont les configurations minimales suffisantes alors que $R_1 = \alpha_1(i_1)$ est une configuration redondante. En effet, R_1 apparaît déjà dans C_{12} et C_{13} Ou un peut les déduire facilement. Ainsi, on obtient :

$$\sum \sum \sum t(i_1, i_2, i_3) \text{Log } t^*(i_1, i_2, i_3) = N\alpha + t(i_1, i_2, +)C_{12} + (i_1, +i_3)C_{13} \quad (\text{E 4.23c}).$$

pour le modèle non saturé (E 4.23a).

D'une manière générale, soit T un tableau p-dimensionnel avec $t(i_\omega)$ l'observation a l'intérieur d'une cellule élémentaire (i_ω) Alors pour un modèle non saturé :

$$\text{Log } t^*(i_\omega) = \alpha_0 + \sum C_i - \sum R_i \quad (\text{E4.24})$$

où $\sum C_i$ et $\sum R_i$ représentent respectivement la somme des configurations minimales suffisantes et la somme des configurations redondantes.

Au-delà de cette notion de configuration minimale suffisante, il importe de préciser deux résultats de Birch(1963) indispensables pour l'estimation:

- « Si $t(i_\omega)$ est un élément de l'ensemble des statistiques suffisantes, l'estimateur du maximum de vraisemblance de $t(i_\omega)$ de $t^*(i_\omega)$ est $t(i_\omega)$ »

- « Il y a une solution positive $t^*(i_w)$ pour chaque cellule du tableau et cette solution satisfait $\sum t^*(i_w) = \sum t(i_w)$ ».

Les résultats de Birch appliqués à C_{12} et C_{13} comme configurations minimales suffisantes, vérifient :

$$\begin{aligned} \hat{t}^*(i_1, i_2, +) &= t(i_1, i_2, +) \\ \hat{t}^*(i_1, +, i_2) &= t(i_1, +, i_2) \end{aligned} \quad (E4.25)$$

Ces éléments préliminaires étant explorés, nous passons à la phase d'estimation proprement dite.

4.2.2. Estimation fondée sur le principe du maximum de vraisemblance.

On rappelle que le principe de cette estimation consiste à rechercher l'ensemble des paramètres qui permettent de maximiser une fonction de vraisemblance, laquelle est obtenue ici lorsque les observations du tableau suivent l'un des processus d'échantillonnage développés en 3.3. Comme on l'a déjà précisé, cette procédure comporte deux approches : la démarche directe et la démarche indirecte

4.2.2.1. Estimation fondée sur la méthode directe.

La méthode directe est celle dans laquelle les conditions de 1^{er} ordre sont remplies et ont des solutions. Le modèle (E 4.23a) peut d'ailleurs être estimé par la méthode directe. Les configurations minimales suffisantes de ce modèle sont C_{12} et C_{13} . De (E 4.23a), on déduit :

$$t^*(i_1, i_2, i_3) = \text{EXP} \left[\alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \alpha_3(i_3) + \alpha_{12}(i_1, i_2) + \alpha_{13}(i_1, i_3) \right]$$

Compte tenu des résultats de Birch, on a

$$t^*(i_1, i_2, +) = \sum_{i_3} \text{EXP} \left[\alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \alpha_3(i_3) + \alpha_{12}(i_1, i_2) + \alpha_{13}(i_1, i_3) \right]$$

$$= \text{EXP} \left[\alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \alpha_3(i_3) \right] \sum_{i_2} \text{EXP} \left[\alpha_2(i_2) + \alpha_{12}(i_1, i_2) \right]$$

De même

$$t^*(i_1, +, i_3) = \sum_{i_3} \text{EXP} \left[\alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \alpha_3(i_3) + \alpha_{12}(i_1, i_2) + \alpha_{13}(i_1, i_3) \right]$$

$$= \text{EXP} \left[\alpha_0 + \alpha_1(i_1) + \alpha_3(i_3) + \alpha_{13}(i_1, i_3) \right] \sum_{i_2} \text{EXP} \left[\alpha_2(i_2) + \alpha_{12}(i_1, i_2) \right]$$

$$\hat{t}^*(i_1, +, +) = \text{EXP} \left[\alpha_0 + \alpha_1(i_1) \right] \sum \sum \text{EXP} \left[\alpha_2(i_2) + \alpha_3(i_3) + \alpha_{12}(i_1, i_2) + \alpha_{13}(i_1, i_3) \right]$$

En multipliant (E 4.26) par (E 4.27) et en divisant le tout par (E4.28), on obtient l'estimateur du maximum de vraisemblance suivant :

$$\bar{t}^*(i_1, i_2, i_3) = \frac{t^*(i_1, i_2, +) t^*(i_1, +, i_3)}{t^*(i_1, +, +)} = \frac{t(i_1, i_2, +) t(i_1, +, i_3)}{t(i_1, +, +)} \quad (\text{E 4.29})$$

Cette approche directe ne peut être appliquée à toutes les configurations suffisantes. Ainsi, du modèle (E2.12), si $\alpha_{123}(i_1, i_2, i_3) = 0$; les configurations minimales suffisantes seront C_{12} , C_{13} , C_{23} . L'application de la méthode directe donnera :

$$\bar{t}^*(i_1, i_2, i_3) = \frac{t(i_1, i_2, +) t(i_1, +, i_3) t(+, i_2, i_3)}{t(i_1, +, +) t(+, i_2, +) t(+, +, i_3)} t(+, +, +)$$

L'utilisation de la méthode indirecte sera donc nécessaire pour estimer le modèle auquel sont associées les configurations C_{12} , C_{13} et C_{23} .

4.2.2.2. Estimation fondée sur la méthode indirecte.

Lorsqu'il n'est pas possible d'appliquer l'approche directe dans la procédure

d'estimation, on peut faire appel à la méthode indirecte, notamment la procédure itérative. Nous avons rappelé (Cf. chap. 3) que deux algorithmes sont généralement utilisés dans cette approche -l'algorithme d'ajustement proportionnel itératif (API) et l'algorithme de Newton-Raphson. Il est important de rappeler encore que ces algorithmes, dans leur logique, empruntent des chemins totalement opposés. En effet, réécrivons le modèle Log-linéaire sous une autre forme, c'est-à-dire :

$$\text{Log } t^* = \mathbf{t} \alpha \quad (\text{E431})$$

où \mathbf{t} est une matrice de "désignation" (Design Matrix) connue et α le vecteur de paramètres ; $t^* = E(t)$ est la valeur estimée de t (t^* et t étant des vecteurs bien entendu).

L'utilisation de l'algorithme d'ajustement proportionnel itératif, conduit dans un premier temps à l'estimation des observations $t(i_0)$, lesquelles permettent de déterminer les valeurs correspondantes.

L'algorithme de Newton-Raphson permet quant à lui d'estimer les valeurs de t^* à partir desquelles son déduit les valeurs estimées de $t(i_0)$. Pour α connu, on tire :

$$t^* = \exp(\mathbf{t} \alpha)$$

A l'exception de l'algorithme d'ajustement proportionnel itératif, toutes les autres méthodes d'estimation emprunteront cette seconde voie. Nous examinerons successivement l'algorithme d'ajustement proportionnel itératif et l'algorithme de Newton-Raphson.

4.2.2.2.1 L'algorithme d'ajustement proportionnel itératif.

La construction de l'algorithme d'ajustement proportionnel itératif a été inspirée de la procédure d'ajustement proportionnel établie par Deming et Stephan (1940), dont le but est de lisser les observations d'un tableau afin que ses marges soient semblables à celles d'un tableau qu'on se donne a priori ou a posteriori (Cf. contraintes internes et externes, voir 3.4.2 et 3.4.3).

Le principe de cet algorithme consiste à partir d'un ensemble de solutions de départ (généralement un tableau uniforme, dont toutes les cellules sont constituées de 1 et de même

dimension que le tableau à estimer) qu'on essaiera d'améliorer au fil des itérations, jusqu'à ce que soient remplies les contraintes de marges (conditions de Birch) ; on dira dans ce cas qu'il y a convergence de l'algorithme. L'algorithme d'ajustement proportionnel itératif converge toujours et on l'applique surtout pour l'estimation des modèles à structure hiérarchique (ajustement des observations aux statistiques minimales suffisantes).

Le modèle (E 4.12) auquel on supprime le terme $\alpha_{23}(i_2, i_3)$ ne pouvant être estimé par l'approche directe, si on lui applique l'algorithme d'ajustement proportionnel itératif, on aura à chaque cycle de l'itération trois étapes correspondant aux configurations minimales suffisantes C_{12} , C_{13} , C_{23} et se traduisant par :

-étape 0

$$\bar{t}_{(0)}^*(i_1, i_2, i_3) = 1 \quad \cdot \quad i_1 = 1, \dots, I_1 ; i_2 = 1, \dots, I_2 ; i_3 = 1, \dots, I_3$$

- étape 1

$$\bar{t}_{(1)}^*(i_1, i_2, i_3) = \bar{t}_{(0)}^*(i_1, i_2, i_3) \frac{t(i_1, i_2, +)}{\bar{t}_{(0)}^*(i_1, i_2, +)} \quad \text{pour } C_{12}$$

- étape 2

$$\bar{t}_{(2)}^*(i_1, i_2, i_3) = \bar{t}_{(1)}^*(i_1, i_2, i_3) \frac{t(i_1, +, i_3)}{\bar{t}_{(1)}^*(i_1, +, i_3)} \quad \text{pour } C_{13}$$

- étape 3

$$\bar{t}_{(3)}^*(i_1, i_2, i_3) = \bar{t}_{(2)}^*(i_1, i_2, i_3) \frac{t(+, i_2, i_3)}{\bar{t}_{(2)}^*(+, i_2, i_3)} \quad \text{pour } C_{23} \text{ (E 4.32a)}$$

Ces différentes étapes constituent le 1^{er} cycle de l'itération. A partir du 2^{ème} cycle de l'itération, on répète les étapes 1, 2, 3 et il y a convergence si pour un seuil fixé δ par exemple, la valeur estimée entre deux itérations successives est sensiblement identique, c'est-à-dire qu'au seuil δ on a;

$$\left| \bar{t}_{(v)}^*(i_1, i_2, i_3) - \bar{t}_{(v-1)}^*(i_1, i_2, i_3) \right| < \delta \quad (\delta > 0)$$

$\bar{t}_{(v)}^*(i_1, i_2, i_3)$ est l'estimateur du maximum de vraisemblance de $t(i_1, i_2, i_3)$ à l'itération v ($v > 0$).

D'une manière générale, si on a un modèle à k configurations minimales suffisantes C_{ω_s} (ω représente les marges; $s = 1, \dots, k$), il existe k étapes dans chaque itération. En partant de la solution initiale $t_{(0)}^*(i_{\omega_s}) = 1$, le résultat après v itérations est:

$$\bar{t}_{(sv+1)}^*(i_{\omega_s}) = \bar{t}_{(sv)}^*(i_{\omega_s}) \frac{t(i_{\omega_s})}{\bar{t}_{(sv)}^*(i_{\omega_s})} \quad (\text{E4.32b})$$

Remarques :

- l'application de l'algorithme d'ajustement proportionnel itératif quand l'approche directe est possible entraîne la convergence de l'algorithme dès la première itération et l'estimateur obtenu par les deux approches est identique.

- la solution de départ est supposée être un tableau dont tous les éléments ne sont que des 1. Cependant, ainsi que le font remarquer Bishop, Fienberg et Holland(1975), n'importe quel tableau peut être choisi comme solution de base, la seule conséquence majeure étant la vitesse de convergence de l'algorithme qui sera plus lente dans le second cas que dans le premier. Nous quittons le cadre de cet algorithme pour voir le mécanisme de fonctionnement de celui de Newton-Raphson.

4.2.2.2.2 L'algorithme d'ajustement de Newton-Raphson.

Il s'agit d'un algorithme couramment utilisé pour la résolution des équations non linéaires. Néanmoins, son utilisation pour l'estimation du modèle Log-linéaire est aussi possible. On suppose que les observations du tableau à estimer par un tel modèle suivent tous un des processus d'échantillonnage décrits en 3.3 et dont le logarithme de la fonction de vraisemblance sous sa forme générale est :

$$\text{Log } L(\alpha_\omega(i_\omega)) = \sum_{i_\omega} t(i_\omega) \text{Log } t^*(i_\omega) - \sum_{i_\omega} t^*(i_\omega) + H(t(i_\omega)) \quad (\text{E 4.33})$$

où $t^*(i_\omega) = E(t(i_\omega))$ est une fonction de $\alpha_\omega(i_\omega)$, et $H(t(i_\omega))$ un terme constant.

Comme $\text{Log } t^*(i_\omega) = \sum_{\omega=0} \alpha_\omega(i_\omega) \theta_\omega(i_\omega)$ (E4.33) devient :

$$\text{Log } \alpha_\omega(i_\omega) = \sum_{i_\omega} t^*(i_\omega) \left[\sum_{\omega=0} \alpha_\omega(i_\omega) \theta_\omega(i_\omega) \right] - \sum_{\omega} \text{EXP} \left[\sum_{\omega=0} \alpha_\omega(i_\omega) \theta_\omega(i_\omega) \right]$$

où bien entendu, $\theta_\omega(i_\omega)$ représente les éléments de la "design-matrix"

Posons $\rho = \text{Log } L(\alpha_\omega(i_\omega))$

$$\frac{\partial \rho}{\partial \alpha_\omega(i_\omega)} = \sum_{i_\omega} t(i_\omega) \theta_\omega(i_\omega) - \sum_{\omega=0} \alpha_\omega(i_\omega) \sum_{\omega} \text{EXP} \left[\sum_{\omega=0} \alpha_\omega(i_\omega) \theta_\omega(i_\omega) \right] \quad (\text{E 4.35})$$

ce qui implique:

$$\sum [t(i_\omega) - t^*(i_\omega)] \theta_\omega(i_\omega) = 0$$

Il faut donc trouver une solution à (E 4.35). La procédure à effectuer sera par itération et pour cela, en se donnant comme solutions initiales les valeurs $\alpha_\omega^{(0)}(i_\omega)$ et on cherche le plan

tangent en $\alpha_\omega^{(0)}(i_\omega)$ à la fonction ρ , $\alpha_\omega(i_\omega) = \frac{\partial \rho}{\partial \alpha_\omega(i_\omega)}$. Ce plan sera d'équation:

$$\frac{\partial \rho}{\partial \alpha_\omega(i_\omega)} (\alpha_\omega^{(0)}(i_\omega)) + \frac{\partial^2 \rho}{\partial \alpha_\omega(i_\omega) \partial \alpha_\omega(i_\omega)} (\alpha_\omega^{(0)}(i_\omega)) [\alpha_\omega(i_\omega) - \alpha_\omega^{(0)}(i_\omega)] \quad (\text{E4.36})$$

Ainsi, ρ , $\alpha_\omega(i_\omega) = 0$ peut être rapprochée par (E4.36) d'où:

$$\frac{\partial \rho}{\partial \alpha_\omega(i_\omega)} (\alpha_\omega^{(0)}(i_\omega)) + \frac{\partial^2 \rho}{\partial \alpha_\omega(i_\omega) \partial \alpha_\omega(i_\omega)} (\alpha_\omega^{(0)}(i_\omega)) [\alpha_\omega(i_\omega) - \alpha_\omega^{(0)}(i_\omega)]$$

On peut en déduire :

$$\alpha_{\omega}^{(1)}(i_{\omega}) = \alpha_{\omega}^{(0)}(i_{\omega}) - \left[\frac{\partial^2 \rho}{\partial \alpha_{\omega}(i_{\omega}) \partial \alpha_{\omega}(i_{\omega})} (\alpha_{\omega}^{(0)}(i_{\omega})) \right]^{-1} \frac{\partial \rho}{\partial \alpha_{\omega}(i_{\omega})} (\alpha_{\omega}^{(0)}(i_{\omega}))$$

En opérant ainsi, on a fait finalement au bout de s itérations :

$$\alpha_{\omega}^{(s)}(i_{\omega}) = \alpha_{\omega}^{(s-1)}(i_{\omega}) - \left[\frac{\partial^2 \rho}{\partial \alpha_{\omega}(i_{\omega}) \partial \alpha_{\omega}(i_{\omega})} (\alpha_{\omega}^{(s-1)}(i_{\omega})) \right]^{-1} \frac{\partial \rho}{\partial \alpha_{\omega}(i_{\omega})} (\alpha_{\omega}^{(s-1)}(i_{\omega}))$$

Il y aura convergence quand on trouvera $\hat{\alpha}_{\omega}(i_{\omega})$ tel que $\frac{\partial \rho}{\partial \alpha_{\omega}(i_{\omega})} \hat{\alpha}_{\omega}(i_{\omega}) = 0$

$$\lim_{s \rightarrow \infty} \alpha_{\omega}^{(s)}(i_{\omega}) = \hat{\alpha}_{\omega}(i_{\omega}) \quad (\text{E4.38})$$

s-----> ∞

$\hat{\alpha}_{\omega}(i_{\omega})$ est l'estimateur du maximum de vraisemblance. Contrairement à l'algorithme d'ajustement proportionnel itératif, l'algorithme de Newton-Raphson ne converge pas toujours (cas d'un tableau où il existe plusieurs cases vides). Haberman(1974) a essayé de l'améliorer en construisant l'algorithme de Newton-Raphson modifié.

Ces principales méthodes d'estimations ne sont pas les seules et il en existe plusieurs autres. Nous en retiendrons quelques unes dans la présentation qui suit, à savoir les méthodes de score, des moindres carrés pondérés et du Khi-deux minimum.

4.2.3 Autres méthodes d'estimation. *

4.2.3.1 Méthode du score

La méthode du score découle de l'algorithme de Newton-Raphson dans lequel on substitue à la matrice hessienne son espérance. Formellement, l'algorithme caractérisant cette méthode aura pour équation de récurrence :

$$\alpha_{\omega}^{(s)}(i_{\omega}) = \alpha_{\omega}^{(s-1)}(i_{\omega}) + E \left[\frac{\partial^2 \rho}{\partial \alpha_{\omega}(i_{\omega}) \partial \alpha_{\omega}(i_{\omega})} (\alpha_{\omega}^{(s-1)}(i_{\omega})) \right]^{-1} \frac{\partial \rho}{\partial \alpha_{\omega}(i_{\omega})} (\alpha_{\omega}^{(s-1)}(i_{\omega})) \quad (\text{E 4.39})$$

Cette méthode est appliquée aux modèles qui ont de bonnes propriétés asymptotiques. En construisant le modèle Log-linéaire sur la base des processus d'échantillonnage poissonnien, multinomial ou produit-multinomial, la méthode du score coïncide avec l'algorithme de Newton-Raphson (Gourieroux, 1980) puisque ces processus ont des propriétés asymptotiques (Bishop, Fienberg et Holland, 1975).

4.2.3.2 La méthode des moindres carrés pondérés.

Cette méthode a été surtout développée par Grizzle, Starmer et Koch(1969) pour son application au modèle d'échantillonnage multinomial. Elle est antérieure aux algorithmes de Newton-Raphson et d'ajustement proportionnel itératif. On rappelle encore que le modèle Log-linéaire peut se mettre sous la forme (E 4.31), c'est-à-dire : $\text{Log } t^* = \mathbf{t} \alpha$.

A cette forme linéaire, il est possible d'appliquer les moindres carrés simples. Mais les valeurs estimées de α risquent d'être biaisées, puisque les observations du tableau ne sont pas toujours homoscédastiques. Comme elles sont en général hétéroscédastiques, on utilise les moindres carrés pondérés pour l'obtention des paramètres. Bien que différente des algorithmes précédents, cette méthode aboutit aux résultats similaires si on estime des tableaux non complexes avec suffisamment d'observations dans chaque cellule (Cf. Kerschner et Chao, 1976). Elle ne donne pas de résultats assez fiables si on a un tableau où il y a plusieurs cellules vides, car se poserait le problème de la régularité de la matrice des variances-covariances (la matrice \mathbf{t} dans le cas présent). Il existe une dernière méthode: celle du Khi-deux minimum.

4.2-3-3 La méthode du Khi-deux minimum.

Par la méthode du Khi-deux minimum, on cherche à estimer un tableau ou un modèle de telle sorte que la distance de Khi-deux entre les observations brutes et les observations

estimées ou encore la distance entre les observations estimées par 2 modèles soit la plus petite. Ainsi, cette méthode consiste à minimiser la statistique :

$$\sum_{i_{\omega}} \frac{[t(i_{\omega}) - t^*(i_{\omega})]^2}{t^*(i_{\omega})} \quad (\text{E4.40})$$

avec $t^*(i_{\omega})$ exprimant, comme fonction de $\alpha_{\omega}(i_{\omega})$ dans le cas du modèle Log-linéaire. L'estimateur est obtenu en écrivant et en résolvant les conditions de 1^{er} ordre et cet estimateur est asymptotiquement équivalent à l'estimateur du maximum de vraisemblance.

En résumé, les différentes méthodes recensées conduisent à l'ajustement d'un tableau donné, ce sur la base d'un modèle caractérisé par des effets. Dans la présentation qui a été faite, nous avons totalement fait abstraction de la qualité du modèle à retenir dans l'estimation. Or, on a vu en introduisant le modèle Log-linéaire, qu'on cherchait à projeter un tableau T dans un sous-espace lisible de manière à ce que cette projection assure une perte d'information minimum. Nous respecterons de manière équivalente ce principe dans l'estimation, ce qui revient à choisir parmi les modèles celui qui ajuste au mieux les données du tableau : c'est le meilleur modèle. Quels sont alors les critères déterminant le "meilleur" modèle ? On ne peut répondre à cette question qu'en examinant les procédures de choix des modèles. Au delà, nous montrerons comment peut se concevoir l'extension du modèle Log-linéaire

4.3. TECHNIQUES DE CHOIX DES MODELES ET EXTENSION DU MODELE LOG-LINEAIRE.

Dans le présent paragraphe, nous serons conduits à effectuer une analyse en deux parties : l'une portant sur les techniques de choix entre des modèles Log-linéaires et l'autre sur l'extension du modèle Log-linéaire.

4.3.1 Techniques de choix entre des modèles Log-linéaires

Il existe plusieurs techniques conduisant au choix du "meilleur" modèle parmi un ensemble de modèles Log-linéaires Bishop(1969), Brown(1976), Goodman(1970, 1971), Kullback et Ku(1968) proposent des approches différentes dont nous rappellerons quelques

éléments. Tout choix de modèle repose sur le critère utilisé pour sa recherche, ce qui implique qu'à critères différents, on ne retrouve pas forcément le même "meilleur" modèle.

Nous présenterons donc différentes méthodes de choix entre modèles, fondées le plus souvent sur les statistiques d'ajustement des modèles. Au passage, nous définirons d'abord ces statistiques avant d'introduire successivement;

- le critère de choix de modèle fondé sur la statistique de G^2 ;
- le critère de choix de modèle fondé sur la procédure Stepwise ;
- le critère de choix de modèle fondé sur les résidus standardisés ;

4.3.1.1 Statistiques de la qualité d'ajustement des modèles

Le choix des modèles Log-linéaires repose principalement sur deux statistiques (il en existe d'autres) ci-après :

- La statistique de Pearson X^2

Elle est définie pour une observation $t(i_\omega)$ estimée par $t^*(i_\omega)$ telle que $E(t(i_\omega)) = t^*(i_\omega)$ par :

$$\sum_{i_\omega} \frac{[t(i_\omega) - t^*(i_\omega)]^2}{t^*(i_\omega)} \quad (\text{E4.41})$$

X^2 suit asymptotiquement une loi de khi-deux à d degrés de liberté (d étant le nombre de degrés de liberté requis pour l'estimation de t^* , vecteur lexicographique des valeurs estimées d'un tableau).

- La statistique de G^2

$$G^2 = 2 \sum t(i_\omega) \text{Log} \frac{t(i_\omega)}{t^*(i_\omega)} \quad (\text{E4.42})$$

qui suit asymptotiquement aussi une loi de Khi-deux à d degrés de liberté également.

Nous utiliserons plus fréquemment la statistique du G^2 , l'un des avantages relatifs à son application résidant dans la souplesse de sa propriété additive (elle peut se décomposer en sous-parties selon les hypothèses formulées), ce qui est très utile lorsque l'on veut mesurer l'importance relative de l'ajout ou de la suppression d'un terme d'un modèle. Ainsi, pour deux modèles M_1 et M_2 tel que M_1 contienne M_2 :

$$G^2 (M_2)= G^2 (M_2/ M_1)+ G^2 (M_1) \quad (E4.43)$$

$G^2 (M_i)$ étant la statistique de G^2 associée au modèle $M_i (i = 1,2)$

$G^2(M_2/ M_1)$ est la statistique conditionnelle du modèle M_2 sachant que le modèle M_1 est vérifié et c'est elle qui permet de mesurer l'importance relative d'un terme ajouté ou supprimé du modèle. En privilégiant cette deuxième statistique, il est normal de regarder comment sur sa base un modèle est choisi.

4.3.1.2 Choix de modèle fondé sur la statistique du G^2 .

La caractéristique essentielle du "meilleur" modèle devra être sa structure simple, c'est-à-dire qu'il doit respecter le principe de la "parcimonie" (modèle à structure non compliquée et constitué d'effets facilement interprétables), qui "réside dans le fait qu'il est souhaitable d'avoir le moins de nombres possibles (les termes dans le cas du modèle Log-linéaire) en vue de concentrer les données à l'aide du modèle". Avec le critère de la statistique du G^2 , nous aborderons la méthode de choix entre deux modèles, les limites de cette statistique, et la définition d'autres statistiques permettant d'aller au-delà de ces limites.

4.3.1.2.1 Principes du choix entre deux modèles.

Soient deux modèles M_1 et M_2 avec M_2 comme cas particulier de $M_1 (M_2 \subset M_1)$ et soient $t^*_1(i_\omega)$ et $t^*_2(i_\omega)$ les valeurs estimées de $t(i_\omega)$ par les modèles M_1 et M_2 respectivement. On utilise pour le choix entre M_1 et M_2 la statistique :

$$G^2 = 2 \sum t(i_\omega) \text{Log} \frac{t^*_1(i_\omega)}{t^*_2(i_\omega)} \quad (E.4.44)$$

En général $G^2(M_2) > G^2(M_1)$ (cette inégalité se vérifie à partir de l'équation (E 4.43)). Cette propriété n'est pas toujours vérifiée avec la statistique X^2 de Pearson. La différence $G^2(M_2) - G^2(M_1)$ (elle mesure l'importance d'un terme ajouté ou supprimé d'un modèle) permet de choisir le "meilleur" modèle entre M_1 et M_2 . Si l'on se donne donc un seuil γ et si la différence entre $G^2(M_2)$ et $G^2(M_1)$ n'est pas significativement différente de zéro, alors le modèle M_2 , plus restreint (c'est le plus parcimonieux) n'entraîne pas une variation substantielle de (E 4.44). Il est donc le meilleur. On rappelle que $G^2(M_2) - G^2(M_1)$ suit un Khi-deux à $d_1 - d_2$ degrés de liberté (d_1 et d_2 sont les degrés de liberté associés à M_1 et M_2). Cependant, cette statistique comme celle de Pearson, a des limites. Nous évoquerons ces limites avant de voir comment elles peuvent être surmontées.

4.3.1.2.2 Limites des statistiques de X^2 et du G^2 .

La valeur prise par la statistique du G^2 ou celle de Pearson dépend non seulement de l'écart existant entre les observations estimées par deux modèles différents, (l'écart peut être aussi entre les observations brutes et celles estimées) mais également de la taille de l'échantillon et ce dernier point peut conduire à des résultats fallacieux au niveau du choix de modèles. Prenons par exemple la statistique du G^2 en (E 4.44). En posant :

$$P_1(i_\omega) = \frac{t_1(i_\omega)}{N} \qquad P_2(i_\omega) = \frac{t_2(i_\omega)}{N}$$

avec $P_1(i_\omega)$ et $P_2(i_\omega)$ les probabilités d'une cellule du tableau par les modèles M_1 et M_2 , on obtient:

$$G^2 = 2N \sum P_1(i_\omega) \text{Log} \frac{P_1(i_\omega)}{P_2(i_\omega)} \qquad (\text{E4.45})$$

Ainsi, lorsque l'on utilise G^2 comme critère de choix des modèles, les tests statistiques qui en découlent sont fonction de la taille de l'échantillon. Dans ce cas, un écart $G^2(M_2) - G^2(M_1)$ sera amplifié par la taille de l'échantillon si elle est surtout grande et la conséquence sera de rejeter le modèle M_2 à tort (risque de première espèce en statistique traditionnelle. Du coup, on peut penser lever l'impasse en diminuant la taille de l'échantillon. Malheureusement en opérant ainsi, d'autres contraintes supplémentaires sont imposées au modèle qu'on teste, si bien qu'on se rapproche d'un modèle à plusieurs restrictions, lequel serait accepté à tort

(risque de deuxième espèce) alors qu'il ne lisse pas au mieux les données initiales du tableau. Dans tous les deux cas -cas de l'importance de la taille de l'échantillon ou lorsque pour un seuil on veut réduire son importance- les statistiques d'ajustement engendrent des erreurs d'appréciation quand à la validité d'un modèle. Afin d'éviter ces erreurs, quelques indices ont été définis et permettent d'annuler l'importance de la taille -"effet taille"- dans les statistiques. Ce sont les indices de Goodman (1971) de D. Bonnet et P. Bentler(1983).

4.3.1.2.3 Indices globaux d'ajustement des modèles

Deux types d'indices sont à distinguer : l'indice d'ajustement normé et l'indice d'ajustement non normé.

Indice d'ajustement normé

Afin de conserver le caractère d'indépendance par rapport à la taille de l'échantillon de la statistique d'ajustement des modèles, Goodman (1971 a défini un indice à normé et tel que, pour deux modèles M_1 et M_2 ($M_2 \subset M_1$), on a:

$$\Delta = \frac{G^2(M_2) - G^2(M_1)}{G^2(M_2)} \quad 0 \leq \Delta \leq 1 \quad (\text{E 4.46})$$

Δ indique l'amélioration relative de l'ajustement de modèle M_1 par rapport au modèle M_2 et ne saurait être interprété comme un coefficient de corrélation, bien que borné par zéro et l'unité.

$-\Delta = 1$ si $G^2(M_1) = 0$, c'est-à-dire si le modèle M_1 ajuste parfaitement les données (c'est le cas général du modèle saturé).

$-\Delta = 0$ si $G^2(M_2) = G^2(M_1)$, ce qui implique une nette amélioration (en termes de paramètres du modèle M_1 , moins parcimonieux que le modèle M_2 .

En partant d'un modèle d'indépendance totale (soit M_2), alors Δ indique la proportion d'association totale du tableau expliquée par le modèle M_1 . Zahn et Fein (1979) pensent que l'indice Δ est indispensable pour la sélection exploratoire des modèles lorsque la taille de l'échantillon est très grande alors que Bentler et Bonnet estiment que cet indice s'applique

également à un tableau à échantillon modeste. Ces deux derniers auteurs construisent un autre indice ressemblant à Δ et qui est non normé.

Indice d'ajustement non normé

Bonnet et Bentler (1983) ont défini l'indice δ par

$$\delta = \frac{G^2(M_2)/d_2 - G^2(M_1)/d_2}{G^2(M_2)/d_2} \quad (\text{E 4.47})$$

Ici et contrairement à Δ , l'indice δ indique l'amélioration relative par degré de liberté du modèle M_1 par rapport au modèle M_2 . Comme Δ , la borne supérieure de δ est atteinte en 1 (cas où M_2 ajuste parfaitement les données du tableau).

Sa borne inférieure n'est plus en 0, puisque la quantité $G^2(M_2)/d_2$ peut être intérieure à $G^2(M_1)/d_1$;

δ est un indice qui peut être complémentaire à Δ , car il a été déjà signalé plus haut que Δ peut être égal à l'unité avec M_1 comme modèle saturé. Dans ce cas, il sera plus préférable d'utiliser l'indice δ pour le choix de modèles. D'ailleurs, la relation suivante existe entre les deux indices

$$\delta = 1 - (1 - \Delta) \frac{d_2}{d_1} \quad \text{d'où } \delta < \Delta \quad (\text{E 4.48})$$

Le critère du G^2 pour le choix de modèle se généralise à la procédure Stepwise.

4.3.1.3 Choix de modèle fondé sur la procédure Stepwise

La procédure Stepwise a été longuement élaborée par Goodman(1971). Il s'agit également d'une procédure fondée sur le G^2 et on peut aussi appliquer à cette procédure, les différents indices servant de correction au G^2 , et qui viennent d'être précédemment présentés. Deux démarches sont possibles dans cette procédure : la "démarche ascendante" et la "démarche descendante". Nous les illustrerons en utilisant 4 modèles hiérarchiques M_1 , M_2 , M_3 et M_4 , susceptibles de servir au lissage d'un tableau tri-dimensionnel.

$$\begin{aligned}
M_1 : \text{Logt}^*(i_1, i_2, i_3) &= \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \alpha_3(i_3) \\
M_2 : \text{Logt}^*(i_1, i_2, i_3) &= \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \alpha_3(i_3) + \alpha_{12}(i_1, i_2) \\
M_3 : \text{Logt}^*(i_1, i_2, i_3) &= \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \alpha_3(i_3) + \alpha_{12}(i_1, i_2) + \alpha_{13}(i_1, i_3) \\
M_4 : \text{Logt}^*(i_1, i_2, i_3) &= \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \alpha_3(i_3) + \alpha_{12}(i_1, i_2) + \alpha_{13}(i_1, i_3) + \alpha_{23}(i_2, i_3)
\end{aligned} \tag{E 4.49}$$

4.3.1.3.1 Approche ascendante

Le point de départ de cette approche est le choix d'un modèle le plus restreint possible. En général, on s'intéresse aux modèles hiérarchiques, ce qui fait que l'on peut retenir comme modèle de départ (hypothèse de départ), le modèle d'indépendance totale (cas du modèle M_1). A l'étape suivante, on obtient un nouveau modèle par ajout d'un terme et on le teste par rapport au précédent pour voir si l'ajout de ce terme améliore le modèle de départ ou pas, suivant développé en (4.1.2.1). La procédure est recommencée avec toutes les combinaisons possibles pouvant être obtenues de manière ascendante, jusqu'à ce que l'on trouve le "meilleur" modèle. Ainsi, en partant de M_1 , on ajoute à celui-ci le terme $\alpha_{12}(i_1, i_2)$ (on a le modèle M_2) et on effectue le test en s'appuyant sur la différence $G^2(M_1) - G^2(M_2)$. Il faut noter qu'avant de passer à M_3 , auront été ajoutés à M_1 tous les termes $\alpha_{12}(i_1, i_2)$, $\alpha_{13}(i_1, i_3)$ ou $\alpha_{23}(i_2, i_3)$ individuellement, afin de déterminer à ce stade quel est le "meilleur" modèle. Si M_1 est le modèle de départ, trois modèles du types M_2 sont possibles à l'étape 2. Il en de même pour les modèles du type M_3 à l'étape 3. A l'étape 4, seul le modèle M_4 est possible. Toutes ces étapes sont recensées par la procédure ascendante pour déterminer le "meilleur" modèle et cette démarche n'est pas la seule ; il en existe une autre, la " démarche descendante", qui est l'opposé de la première.

4.3.1.3.1 Approche ascendante

C'est une approche inverse de la procédure précédente. Elle consiste cette fois-ci à rechercher le "meilleur" modèle par élimination successive d'un ou plusieurs termes d'un modèle quelconque à structure complexe. Si l'on prend ici M_4 comme modèle de départ, on le testera d'abord par rapport à M_3 pour apprécier l'amélioration relative de M_4 quand on supprime le terme $\alpha_{23}(i_2, i_3)$. Toutes les combinaisons possibles nous conduisent au "meilleur" modèle, par manipulation " descendante " et à l'inverse du chemin emprunté dans l'approche

"ascendante". L'objectif est donc de réduire davantage le modèle en s'assurant d'une qualité minimum de l'ajustement obtenu.

Remarques

- quelle que soit l'approche retenue, le "meilleur" modèle reste totalement lié au modèle retenu au départ, d'où son caractère non unique.

- choisir un modèle, c'est tester les hypothèses pour un seuil donné. Le "meilleur" modèle reste donc tributaire du seuil fixé pour son choix.

- même avec un seuil identique, on ne converge pas forcément vers le même modèle en empruntant les deux approches décrites précédemment. Goodman (1971) propose alors la combinaison des deux approches afin d'aboutir à un modèle plus intéressant que le "meilleur" obtenu par l'une des deux approches.

Cette procédure, appliquée à un tableau ayant plusieurs variables, peut engendrer des coûts informatiques assez considérables, qui restent néanmoins modestes en ce qui concerne les tableaux de dimension faible. Peut-on donc l'envisager à grands frais lorsque l'on sait son caractère arbitraire pour la détermination du "meilleur" modèle ?

Une autre méthode de choix de modèle est proposée et elle reste fondée sur les paramètres standardisés,

4.3.1.4. Choix de modèle fondé sur les termes standardisés des paramètres.

Le mécanisme de ce choix est le suivant : à partir d'un modèle Log-linéaire saturé, dont on estime les variances et par conséquent les écarts types associés aux paramètres, on fait le test de Student de ces paramètres (KEITH, LEE , 1977). S'ils sont significativement différents de zéro, ils sont retenus dans le modèle, sinon, ils en sont éliminés. En effectuant ce test, on normalise en même temps les termes d'effets du modèle, d'où l'appellation terme standardisé, qui s'écrit :

$$u_{\omega}(i_{\omega}) \frac{\alpha_{\omega}(i_{\omega})}{\sigma_{\omega}(i_{\omega})} \quad (\text{E 4.50})$$

où $\sigma_{\omega}(i_{\omega})$ est l'écart type du terme $\alpha_{\omega}(i_{\omega})$ et $u_{\omega}(i_{\omega})$ le terme standardisé de $\alpha_{\omega}(i_{\omega})$.

En principe l'élimination des termes non significatifs devrait permettre de constituer le modèle à choisir avec les termes restants. Le test de Student, d'autre part, fait appel à l'hypothèse de normalité alors que nous avons précisé plus loin que le modèle Log-linéaire n'a aucun caractère aléatoire, ce qui met déjà en cause cette méthode. En somme, on peut penser que la méthode fondée sur les termes standardisés des paramètres est plus une étape préliminaire à la sélection des modèles qu'une technique de choix proprement dite. En présence d'un tableau de très grande dimension, il devient difficile de trouver un modèle de départ pouvant être amélioré par les approches "ascendante" ou "descendante" et c'est ici que peuvent servir les termes standardisés pour l'obtention du modèle de base.

Cette première approche du modèle Log-linéaire a eu pour objet de mettre l'accent sur sa formulation de base, avec au passage le développement des techniques d'estimation et de choix de modèle. Il est néanmoins possible de dépasser le cadre traditionnel que constitue cette approche et exploiter le modèle Log-linéaire à d'autres fins, en particulier son utilisation pour l'allocation de l'information totale d'un tableau quelconque entre les différentes marges indépendantes de ce tableau ou encore entre les marges entrant en compte dans le modèle de lissage retenu. C'est ce que nous montrerons au paragraphe (4.3.2).

4.3.2 Extension du modèle Log-linéaire : son approche informationnelle.

Au-delà du modèle Log-linéaire de base, nous allons développer ici l'approche élaborée par ZIGHERA(1985), et dont l'objet est de permettre le partitionnement de l'information totale contenue dans un tableau entre les marges de ce tableau. Ce partitionnement est en aval de la procédure de centrage des paramètres du modèle -par cette procédure, les paramètres ont une signification alors que dans le modèle de base ils n'en avaient pas- Mais avant d'en arriver, nous aurons rappeler auparavant et successivement l'objet du centrage des paramètres, la notion et les propriétés des résidus, et la procédure de centrage des paramètres proprement dite.

4.3.2.1 Objet du centrage des paramètres.

Revenons au modèle Log-linéaire défini dans sa forme générale en (E 4.9b) et retenons le cas simple d'un modèle non saturé que nous appliquons à un tableau T de dimension 3 et qui est tel que :

$$\text{Log}t^*(i_1, i_2, i_3) = \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \alpha_3(i_3) + \alpha_{12}(i_1, i_2) + \alpha_{13}(i_1, i_3) + \alpha_{23}(i_2, i_3)$$

avec les contraintes

$$\begin{aligned} \sum_{i_1} \alpha_1(i_1) &= \sum_{i_2} \alpha_2(i_2) = \sum_{i_3} \alpha_3(i_3) = 0 \\ \sum_{i_2} \alpha_{23}(i_2, i_3) &= \sum_{i_3} \alpha_{23}(i_2, i_3) = 0 \end{aligned} \quad (\text{E 4.51})$$

Le modèle (E 4.51) a $I_2 I_3 + I_1 - 1$ degrés de contraintes.

Soit α le vecteur des paramètres de ce modèle ; α est donc un vecteur à $I_2 I_3 + I_1 - 1$ composantes. Celles-ci ne sont pas toutes linéairement indépendantes puisqu'elles sont liées aux contraintes de marges, lesquelles sont également liées par une relation de dépendance (Cf. marges impliquantes, marges impliquées (Cf. 3.2.3)). D'autre part, les contraintes de (E 4.51) comme toutes les autres qui ont jusqu'ici été imposées au modèle Log-linéaire sont arbitraires, bien que permettant d'obtenir un vecteur α dans la base de la décomposition du modèle (Cf. Gourieroux (1980)). Le vecteur $\alpha \in Y$ où Y est un hyperplan) et plusieurs combinaisons possibles des composantes de α peuvent conduire à l'estimation de la valeur $t(i_1, i_2, i_3)$ du tableau T , lissé par le modèle (E 4.51). Cette situation ne pose a priori aucun problème si ce modèle ne se limite qu'à la prospection des relations pouvant exister entre les différentes variables du tableau. L'interprétation même des paramètres n'a aucune signification concrète. En effet, si on se limite pour l'instant aux effets de paramètres du 1^{er} ordre, on ne pourra conclure, suivant la grandeur de ces effets, qu'à la sur-représentativité (lorsque les effets sont positifs) ou la sous-représentativité (les effets sont négatifs) des modalités auxquelles ils sont associés. Il n'est donc pas possible de trouver une échelle structurée (ordonnée) des effets de paramètres dans leur conception de base. Par conséquent, si l'on veut dépasser le cadre purement exploratoire des paramètres comme nous venons de le décrire et s'ouvrir à d'autres analyses, notamment rechercher par exemple comment se positionnent les différentes modalités des variables constituant un tableau afin d'en tirer des conclusions fiables et pertinentes, voire effectuer des comparaisons non erronées, il serait souhaitable de trouver un univers approprié pour de telles analyses. Cet univers sera défini en procédant au **centrage des paramètres**, à la base de la décomposition de l'information totale d'un tableau, c'est-à-dire son partitionnement entre les différentes

marges indépendantes qui entrent en compte dans le modèle retenu pour le lissage du tableau T. Auparavant, un rappel aura été fait sur la notion et les propriétés des résidus, lesquels interviennent dans le processus de centrage.

4.3.2.2 Notion et propriétés des résidus.

Les résidus couramment utilisés en statistique traditionnelle sont différemment définis en statistique multiplicative où ils respectent la nature multiplicative (écart relatif) de l'approche. En effet, ici, il s'agit de l'écart entre le logarithme de l'observé et le logarithme de l'estimé, c'est-à-dire le logarithme du rapport (Observé/Estimé).

Soit donc un tableau T 3-dimensionnel avec $t(i_1, i_2, i_3)$ la valeur de l'observation à l'intérieur d'une cellule de T ; soit également $r_{123}(i_1, i_2, i_3)$ le résidu d'une cellule de T.

Alors,

$$\begin{aligned} r_{123}(i_1, i_2, i_3) &= \text{Log}t(i_1, i_2, i_3) - \text{Log}t^*(i_1, i_2, i_3) \\ &= \text{Log} \frac{t(i_1, i_2, i_3)}{t^*(i_1, i_2, i_3)} \end{aligned} \quad (\text{E 4.52})$$

De cette relation, on déduit :

$$\text{EXP}(r_{123}(i_1, i_2, i_3)) = \frac{t(i_1, i_2, i_3)}{t^*(i_1, i_2, i_3)} \quad (\text{E 4.53})$$

$$\text{d'où } t(i_1, i_2, i_3) \text{EXP}(-r_{123}(i_1, i_2, i_3)) = t^*(i_1, i_2, i_3)$$

Compte tenu des contraintes de BIRCH (1963) et l'implication des marges, on a par exemple :

$$\sum_{i_2} \sum_{i_3} t(i_1, i_2, i_3) \text{EXP}(-r_{123}(i_1, i_2, i_3)) = \sum_{i_2} \sum_{i_3} t^*(i_1, i_2, i_3) = t^*(i_1, +, +) \quad (\text{E 4.54})$$

$$\text{d'où } \sum_{i_2} \sum_{i_3} \frac{t(i_1, i_2, i_3)}{t^*(i_1, i_2, i_3)} \text{EXP}(-r_{123}(i_1, i_2, i_3)) = 1 \quad (\text{E 4.55a})$$

Supposons pour simplifier que $t^*(+, +, +) = t(+, +, +) = 1$. Alors,

$$\sum_{i_2} \sum_{i_3} t^*(i_1, i_2, i_3) \text{EXP}(-r_{123}(i_1, i_2, i_3)) = t(+, +, +) = 1 \quad (\text{E 4.55b})$$

De manière analogue à (E 4.55a), on a :

$$\sum_{i_2} \sum_{i_3} \frac{t(i_1, i_2, i_3)}{\text{EXP}(r_{123}(i_1, i_2, i_3))} = 1 \quad (\text{E 4.55c})$$

Deux conclusions sont tirées des équations (E 4.54) à (E 4.55c) :

- la somme pondérée des résidus respecte les marges ;
- la moyenne harmonique des résidus exponentialisés et pondérés est égale à l'unité.

Il est également possible de calculer les moyennes (pondérées) des résidus :

$$r_{123}(+, +, i_3) = \sum_{i_2} \sum_{i_3} \frac{t(i_1, i_2, i_3)}{t^*(i_1, i_2, i_3)} r_{123}(i_1, i_2, i_3)$$

est la moyenne arithmétique pondérée des résidus.

Ces notions et propriétés rappelées, il se pose maintenant le problème de la procédure de centrage.

4.3.2.3 Procédure de centrage des paramètres.

Supposons retenu le modèle défini en (E 4.51) et servant à lisser le tableau T. On peut écrire :

$$\text{Log} t^*(i_1, i_2, i_3) = \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \alpha_3(i_3) + \alpha_{12}(i_1, i_2) + \alpha_{13}(i_1, i_3) + \alpha_{23}(i_2, i_3) \quad (\text{E 4.57})$$

On rappelle que le terme $\alpha_{23}(i_2, i_3)$ mesure l'effet de structure de la facette formée par les marges $M(i_2)$ et $M(i_3)$ du tableau. Ce terme comprend les effets provenant de $M(i_2)$ et $M(i_3)$ et les effets propres à $M(i_2, i_3)$. Il doit donc falloir "essorer" $\alpha_{23}(i_2, i_3)$ afin de conserver uniquement les effets qui lui sont propres. Comme $\alpha_{23}(i_2, i_3)$ peut s'exprimer comme combinaison linéaire d'un ensemble d'autres paramètres $\alpha_{23}(i_2, i_3) \in Y$, alors :

$$\alpha_{23}(i_2 i_3) = \beta_0 + \beta_2(i_2) + \beta_3(i_3) + \beta_{23}(i_2, i_3) \quad (\text{E 4.58})$$

La modélisation de $EXP(\alpha_{23}(i_2, i_3))$ donne :

$$\text{Log}(EXP\alpha_{23}(i_2 i_3)) = \beta_2(i_2) + \beta_3(i_3) + \zeta_{23}(i_2, i_3) \quad (\text{E 4.59})$$

où $\zeta_{23}(i_2, i_3)$ est le terme centré de $\alpha_{23}(i_2, i_3)$.

Une fois que $\alpha_{23}(i_2, i_3)$ est "essoré", on retrouve facilement les effets dus aux marges $M(i_2)$ et $M(i_3)$ corrigés d'une partie des effets qui étaient intégrés dans $M(i_2, i_3)$.

Désignons par $\zeta_1(i_1)$, $\zeta_2(i_2)$ et $\zeta_3(i_3)$ les paramètres centrés de $\alpha_1(i_1)$, $\alpha_2(i_2)$ et $\alpha_3(i_3)$ respectivement. Si l'on pose :

$$\begin{aligned} \rho_1(i_1) &= \alpha_1(i_1) \\ \rho_2(i_2) &= \alpha_2(i_2) + \beta_2(i_2) \\ \rho_3(i_3) &= \alpha_3(i_3) + \beta_3(i_3) \end{aligned} \quad (\text{E 4.60})$$

Alors :

$$\begin{aligned} \zeta_1(i_1) &= \rho_1(i_1) + \text{Log} [\sum t(i_1, +, +) \text{EXP}(-\rho_1(i_1))] \\ \zeta_2(i_2) &= \rho_2(i_2) + \text{Log} [\sum t(+, i_2, +) \text{EXP}(-\rho_2(i_2))] \\ \zeta_3(i_3) &= \rho_3(i_3) + \text{Log} [\sum t(+, +, i_3) \text{EXP}(-\rho_3(i_3))] \end{aligned} \quad (\text{E 4.61})$$

Lorsque ces paramètres sont ainsi centrés, les contraintes que respectent $\zeta_1(i_1)$, $\zeta_2(i_2)$, $\zeta_3(i_3)$ et $\zeta_{23}(i_2, i_3)$ ne sont plus comme celles imposées en (E4.51) (c'est-à-dire les contraintes ANOVA), mais plutôt les suivantes :

$$\begin{aligned} \sum \frac{t(i_1, +, +)}{\text{EXP}(\xi_1(i_1))} &= 1 \\ \sum \frac{t(+, i_2, +)}{\text{EXP}(\xi_2(i_2))} &= 1 \\ \sum \frac{t(+, +, i_3)}{\text{EXP}(\xi_3(i_3))} &= 1 \end{aligned} \quad (\text{E4.62})$$

$$\sum_{i_2} \sum_{i_3} \frac{t(+, i_2, i_3)}{\text{EXP}(\zeta_{23}(i_2, i_3))} = 1$$

Ces contraintes indiquent que les moyennes harmoniques des différents paramètres centrés, exponentialisés et pondérés par les marges associées sont égales à l'unité. Nous utiliserons ces paramètres centrés pour décomposer l'information du tableau entre les marges contraignantes d'un tableau.

4.3.2.4 Décomposition de l'information entre les marges indépendantes d'un tableau.

Avant de décomposer l'information entre les marges contraignantes d'un tableau T, nous allons rappeler comment dans ce tableau, se quantifie l'information définie en début du chapitre.

Dans le cas d'un tableau de dimension 3, où les variables V_1 , V_2 et V_3 du tableau ont I_1 , I_2 et I_3 modalités respectivement, l'information contenue dans le tableau T, notée $N(T)$ est telle que:

$$N(T) = \sum \sum \sum \frac{t(i_1, i_2, i_3)}{t(+, +, +)} \text{Log} I_1 I_2 I_3 \frac{t(i_1, i_2, i_3)}{t^*(+, +, +)} \quad (\text{E4.63})$$

En simplifiant (E4.63) à un facteur d'échelle près, on trouve:

$$N(T) = \sum \sum \sum \frac{t(i_1, i_2, i_3)}{t(+, +, +)} \text{Log} \frac{t(i_1, i_2, i_3)}{t^*(+, +, +)} \quad (\text{E4.64})$$

$$\text{Or, } \text{Log} t(i_1, i_2, i_3) = \text{Log} t^*(i_1, i_2, i_3) + r_{123}(i_1, i_2, i_3) \quad (\text{E 4.65})$$

L'équation (E 4.65) découle de (E 4.52). En la substituant dans (E 4.64), compte tenu du modèle (E 4.59), on optimise l'information du tableau qui se décompose en sous-parties indépendantes, d'où;

$$N(T) = \sum \sum \sum \frac{t(i_1, i_2, i_3)}{t(+, +, +)} [\zeta_1(i_1) + \zeta_2(i_2) + \zeta_3(i_3) + \zeta_{23}(i_2, i_3) + r_{123}(i_1, i_2, i_3)]$$

$$\begin{aligned}
&= \sum \frac{t(i_1, +, +)}{t(+, +, +)} \zeta_1(i_1) + \sum \frac{t(+, +, i_2)}{t(+, +, +)} \zeta_2(i_2) + \sum \frac{t(+, +, i_3)}{t(+, +, +)} \zeta_3(i_3) \\
&\quad + \sum \sum \frac{t(i_1, i_2, i_3)}{t(+, +, +)} \zeta_{23}(i_2, i_3) + \sum \sum \sum \frac{t(i_1, i_2, i_3)}{t(+, +, +)} r_{123}(i_1, i_2, i_3) \text{ (E 4.66)}.
\end{aligned}$$

$\sum \frac{t(i_1, +, +)}{t(+, +, +)} \zeta_1(i_1)$ est la contribution absolue de la première variable à l'information totale du tableau.

$\sum \frac{t(+, +, i_2)}{t(+, +, +)} \zeta_2(i_2)$ est la contribution absolue de la deuxième variable à l'information totale du tableau.

$\sum \frac{t(+, +, i_3)}{t(+, +, +)} \zeta_3(i_3)$ est la contribution absolue de la troisième variable à l'information totale du tableau.

$\sum \sum \frac{t(i_1, i_2, i_3)}{t(+, +, +)} \zeta_{23}(i_2, i_3)$ est la contribution absolue des variables V_2 et V_3 à l'information totale du tableau.

CONCLUSION

L'étude du modèle Log-linéaire nous a conduit à, l'exploration de ses deux approches différentes :

- la formulation de base du modèle Log-linéaire ;
- l'extension du modèle Log-linéaire ;

La formulation de base du modèle Log-linéaire a pour but essentiel d'élucider et de décrypter de multitudes structures latentes existant dans un tableau. C'est pourquoi les paramètres du modèle n'ont aucun caractère significatif et de surcroît, ne peuvent être interprétés, bien que WILSON (1979) estime que "leur interprétation ne se justifie que dans certaines circonstances spéciales et non universelles".

L'extension du modèle Log-linéaire permet dans un premier temps de corriger les différents paramètres du modèle avant de les utiliser pour allouer l'information du tableau entre les marges indépendantes. Ainsi, sont détectées les marges (sous-tableaux), qui apportent une information importante au tableau, c'est-à-dire les marges ayant une contribution à l'information forte. Sur la base de ce critère, il peut être possible de vérifier si les marges qui entrent en compte dans le "meilleur" modèle sont celles ayant une contribution à l'information forte, ce qui serait un élément d'évaluation du "meilleur" modèle. Il se pose encore le problème d'arbitraire, car ce critère dépendra du seuil et on se demande à partir de quel seuil la contribution à l'information est forte.

Il subsiste néanmoins un biais dans les deux conceptions du modèle Log-linéaire : celui induit par la nature des variables. Jusqu'ici, toutes les variables du tableau d'étude sont supposées tout simplement nominales. Or, comme nous l'avons signalé au chapitre 3, les données sur lesquelles nous travaillons sont qualitatives, les unes étant nominales, les autres ordinales. En faisant donc table rase sur la nature ordonnée de certaines variables du tableau, on perd de l'information si ces variables existent dans un tableau à étudier. D'autre part, la formulation de base est incapable de nous restituer une telle information. Il est donc indispensable, de rechercher dans le cadre du modèle Log-linéaire, une formulation qui intègre à la fois la nature qualitative et ordonnée des variables.

CHAPITRE V

MODELISATION LOG-LINEAIRE DES TABLEAUX DE CONTINGENCE AYANT DES MODALITES ORDINALES

5.0 INTRODUCTION

Un certain nombre de modèles pouvant servir à l'analyse des données statistiques ont été abordés au chapitre 3 (cf. 3.5). Afin que toute analyse faite par le biais d'un quelconque modèle conduise à des résultats appréciables, il est important que la nature des données, c'est-à-dire leurs caractéristiques principales et/ou leurs spécificités en soit prise en compte de manière explicite ou implicite. On peut donc constater qu'un modèle de régression s'applique très parfaitement aux données de nature quantitative tandis qu'un modèle de lissage par double polynômes orthogonaux s'adapte aux données chronologiques où il est question par exemple de ressortir le degré de représentation d'une génération (cohorte) donnée à des périodes différentes et successives, pour ne citer que ces deux exemples.

Le modèle Log-linéaire comme on l'a vu, est un outil de traitement des variables qualitatives. Les deux composantes essentielles de celles-ci étant les variables ordinales et nominales. Cependant, les études ayant exploité ce modèle à l'origine ont traité indifféremment aussi bien les variables nominales que les variables ordinales (BISCHOP, FIENBERG et HOLLAND, 1975). Comme l'ordre a un sens, toute abstraction faite de cette caractéristique en présence des variables ordinales entraîne une sous-exploitation de "l'information contenue dans les données, puisque tous les aspects de ces données ne sont pas pris en compte".

Plusieurs études récentes ont tenté d'intégrer au modèle Log-linéaire la "dimension" ordre. On pense notamment aux travaux de BOCK et YATES (1973), FIENBERG (1979), GOODMAN(1979, 1981), ANDERSEN (1980), CLIFFORD et CLOGG (1982), AGRIESTI (1983) et RICHARD BREEN (1984). En général, cette intégration se fait par assignation des scores (il ne s'agit pas du système de pondération au sens strict) aux modalités des variables ordinales - cette assignation est arbitraire et se fait suivant l'échelle dans laquelle sont classées les modalités - Bien évidemment la forme structurelle du modèle Log-linéaire qui en découle n'est pas identique à celle que l'on a rencontré jusqu'ici et on l'appellera "modèle d'association", le rapprochement bien que dans une optique différente pouvant se faire avec la mesure d'association de D.HUTTCHINSON (1985) ou de McCULLAGH (1930) (on rappelle, pour se fixer les idées qu'un modèle d'association s'applique à un tableau ayant des variables ordinales).

L'opération d'assignation arbitraire des scores reste une opération totalement en dehors des données et du modèle. Lorsqu'ils sont estimés à l'intérieur du modèle, on arrive non plus au modèle Log-linéaire, mais au modèle Log-multiplicatif (RICHARD BREEN (1984) ANDERSEN(1980)) ou Log non linéaire (GOODMAN (979)).

L'objet du présent chapitre est de mieux préciser la formulation des modèles d'association tant bien dans une logique multiplicative que dans une approche qui ne l'est pas fondamentalement, mais qui par extension peut être considérée comme un aspect du modèle Log-linéaire.

Ce chapitre comporte deux grandes sections :

La première est celle servant de cadre à la description des modèles d'association, conçus sous la forme Log-linéaire. Il s'agit des modèles Log-linéaires qui prennent en compte l'information relative à l'ordre de certaines variables du tableau telle que nous l'avons indiqué plus haut. On pourra ainsi développer la formulation de ces modèles successivement dans un tableau de dimension 2 et de dimension 3, ce dernier cas nous permettant d'envisager une formulation à l'ordre p ($p > 3$).

La deuxième section quant à elle aborde d'autres types de modèles pouvant également servir au traitement des tableaux ayant des variables ordinales et parmi lesquels le modèle logit (5.2) qui peut se présenter comme un cas particulier du modèle Log-linéaire mais avec une structure plus parcimonieuse, la paramétrisation par doubles polynômes orthogonaux dont la décomposition est aussi Log-linéaire et se démarquant des modèles précédents par l'unicité des paramètres, et enfin le modèle Log-non linéaire, qui diffère du modèle Log-linéaire par sa caractéristique principale qui est la non connaissance des scores liés aux modalités ordonnées. Dans cette section, il sera ébauché une voie visant à utiliser le modèle Log-non linéaire dans une approche de la méthode de l'analyse des correspondances.

5.1 MESURE D'ASSOCIATION ET FORME STRUCTURELLE DU MODELE LOG-LINEAIRE APPLIQUE A UN TABLEAU A VARIABLES ORDINALES.

La forme générale du modèle Log-linéaire permet d'effectuer la reparamétrisation d'un tableau de contingence à travers des termes d'effets de paramètres. L'existence d'une échelle ordonnée entre les modalités entraînera une modification de certains termes d'effets de façon à

permettre l'inclusion de l'information relative à la nature ordinale des variables dans le modèle.

L'association (ne pas confondre avec le modèle d'association) est une mesure de la corrélation entre deux variables. Cette corrélation peut être soit linéaire, soit non linéaire et c'est suivant la nature de cette corrélation que dépendra la forme structurelle des modèles d'association que nous examinerons.

Dans ce paragraphe, sera étudiée la formulation Log-linéaire des modèles d'association. Après avoir défini la mesure générale de l'association (5.1.1), l'analyse sera centrée sur les modèles d'association fondés sur les odd-ratios (5.1.2) lesquels auront un lien, par une certaine transformation, avec la forme Log-linéaire des modèles d'association qui sera décrite en (5.1.3)? de même que ses différentes approches dans les tableaux aussi bien bidimensionnels que tri-dimensionnels.

5.1.1 Mesure générale de l'association

Ainsi que le souligne JAE-ON KIM (1975) "lorsque l'on parle de la mesure d'association, il s'agit d'un indice désignant la mesure du degré dans lequel deux ou plusieurs variables " vont ensemble" dans un certain sens". A la lumière de cette citation, se dégagent deux idées fortes :

- il s'agit avant tout d'un indice et on conclue que l'association dans un tableau p-dimensionnel ($p > 2$) peut être saisie par cet indice. Celui-ci est global puisqu'il n'est pas possible de distinguer les associations partielle, conditionnelle et uniforme (cf. GOODMAN, 1979).

- l'expression "vont ensemble" traduirait l'existence d'une liaison entre les variables, ce qui illustrerait la finalité de cet indice comme élément d'évaluation de la liaison globale entre deux ou plusieurs variables.

On suppose, deux variables V_1 et V_2 et deux modèles M_1 et M_2 les décrivant :

M_1 : il y a indépendance statistique entre V_1 et V_2 ;

M_2 : l'une des variables est liée fonctionnellement à l'autre (la fonction est linéaire ou curvilinéaire etc..

JAE - ON KIM définit son indice, le P.R.E. (Proportional -Reduction-in-Error ou probabilité de réduction d'erreur ") par :

$$\text{P.R.E.} = \frac{\text{Probabilité d'erreur en } (M_1) - \text{probabilité d'erreur en } (M_2)}{\text{Probabilité d'erreur en } (M_1)} \quad (\text{E } 5.1)$$

L'équation (E5.1) décrit la mesure relative de l'écart d'erreur en situation d'indépendance par rapport à la forme fonctionnelle entre V_1 et V_2 pour M_2 . Le P.R.E. mesure alors l'association pouvant être interprétée comme une déviation relative par rapport à l'indépendance, c'est-à-dire le degré dans lequel la relation entre V_1 et V_2 dévie de l'indépendance statistique pour s'approcher de la forme fonctionnelle en (M_2).

Le calcul du P.R.E. reste néanmoins fondé sur le modèle de régression, puisqu'il postule entre autre une certaine relation fonctionnelle entre V_1 et V_2 . Notre étude se prêtant à des tableaux multicroisés, ayant des données plutôt qualitatives, la régression ne semble pas adaptée à leur traitement (voir 3.1.2), ou du moins, ne cadre pas avec la méthodologie qu'on se propose d'entreprendre dans la présente étude.

Les modèles d'association que nous présenterons, tout en étant des déviations par rapport à l'indépendance (modèle d'indépendance) se démarqueront du fondement qui guide le P.R.E. de JAE-ON KIM. Il s'agira ici de la combinaison de l'analyse structurelle des tableaux de contingence et celle de la régression (DUCAN, 1979). Cette nouvelle approche, qui ressemble à l'analyse de la covariance (LEHMAN, 1959) caractérisera par la suite les modèles d'association (5.1.3) sous leur forme Log-linéaire, laquelle pourra également découler des modèles d'association fondés sur les odd-ratios, et dont nous définissons quelques-uns à partir d'un tableau de dimension 2.

5.1.2. Modèles d'association fondés sur les odd-ratios

En général, les odd-ratios portent sur un tableau où toutes les variables sont dichotomiques. Cependant, dans un tableau de dimension 2 à variables polytomiques, il est possible de construire autant de odd-ratios que de sous- tableaux 2x2 (c'est le odd-ratio d'ordre 2 et dans le cas par exemple d'un odd-ratio d'ordre 3, on aura autant de odd-ratios que de sous-tableaux 2x2x2, ce principe s'étendant bien entendu au odd-ratio d'ordre p ($p > 3$). En

présence d'un tableau à deux dimensions où V_1 et V_2 sont des variables à I_1 et I_2 modalités ($I_1 > 2$; $I_2 > 2$), il existe :

$$\frac{I_1 I_2 (I_1 - 1)(I_2 - 1)}{4}$$

odd-ratios parmi lesquels certains n'ont pas d'intérêt. Si l'on ne s'en tient qu'aux modalités adjacentes, alors seront retenus $(I_1 - 1)(I_2 - 1)$ odd-ratios permettant de déterminer l'association totale dans ce tableau ;

"l'association locale mesurée par des sous-tableaux formés à partir des lignes (ici les modalités I_1) et les colonnes (ici les modalités I_2) adjacentes du tableau détermine l'association totale de ce tableau" (Goodman, 1979).

$$\theta_{i_2}(i_1, i_2) = \frac{t^*(i_1, i_2) \bullet t^*(i_1 + 1, i_2 + 1)}{t^*(i_1 + 1, i_2) \bullet t^*(i_1, i_2 + 1)}$$

$$i_1 = 1, \dots, I_1$$

$$i_2 = 1, \dots, I_2 \quad (E 5.2)$$

où $\theta_{i_2}(i_1, i_2)$ représente les différents odd-ratios sur la base desquels seront définis les modèles d'association nulle (modèle d'indépendance), d'association uniforme, d'association " ligne", d'association " colonne " ou d'association " ligne-colonne " (on rappelle qu'on se limite pour l'instant à un tableau bidimensionnel).

5.1.2.1 Modèle d'association nulle

Le modèle d'association nulle correspond au modèle d'indépendance. Il se caractérise par :

$$\theta_{i_2}(i_1, i_2) = 1 \quad \begin{array}{l} i_1 = 1, \dots, I_1 - 1 \\ i_2 = 1, \dots, I_2 - 1 \end{array} \quad (E 5.3)$$

Ce modèle ne permet d'estimer aucun paramètre. Il a donc zéro degré de contraintes, soit $(I_1 - 1)(I_2 - 1)$ degrés de liberté. Si l'association existe dans le tableau et qu'elle est la même dans tous les sous-tableaux 2×2 , on a alors la caractéristique d'un modèle d'association uniforme.

5.1.2.2. Modèle d'association uniforme

Le modèle d'association uniforme est décrit par:

$$\theta_{i_2}(i_1, i_2) = \gamma \quad \begin{array}{l} i_1=1, \dots, I_1-1 \\ i_2=1, \dots, I_2-1 \end{array} \quad (E 5.4)$$

Par rapport au modèle d'association nulle, il y a un paramètre à estimer (le paramètre γ). Le modèle (E 5.4) a un degré de contrainte, soit $(I_1 - 1) (I_2 - 1) - 1$ degrés de liberté.

5.1.2.3. Modèle d'association "ligne" (variable V_1)

Le modèle d'association "ligne" prend la forme suivante :

$$\theta_{i_2}(i_1, i_2) = \theta_{i_2}(i_1, +) \quad \begin{array}{l} i_1=1, \dots, I_1-1 \\ i_2=1, \dots, I_2-1 \end{array} \quad (E 5.4)$$

Ce modèle traduit l'abstraction faite aux modalités colonnes (variable V_2) afin d'évaluer uniquement les effets dus à la variable V_1 . Il est similaire au modèle de Simon (1974) où la variable V_2 est ordonnée de telle sorte que l'on calcule les effets de la variable V_1 qui mesurent l'association. Le modèle (E 5.5) a $(I_1 - 1)$ degrés de contraintes soit $(I_1 - 1) (I_2 - 2)$ degrés de liberté et il est utilisé lorsque $I_2 > 2$ ce qui est évident avec les hypothèses de départ. De manière tout à fait symétrique, on détermine également le modèle d'association " colonne".

5.1.2.4. Modèle d'association "colonne"

Le modèle d'association " colonne " sera défini par:

$$\theta_{i_2}(i_1, i_2) = \theta_{i_2}(i_1, +) \quad \begin{array}{l} i_1=1, \dots, I_1-1 \\ i_2=1, \dots, I_2-1 \end{array} \quad (E 5.6)$$

Il a quant à lui I_2-1 degrés de contraintes soit $(I_1 - 2) (I_2 - 1)$ degrés de liberté. Ce modèle comme le précédent ressemble au modèle permettant de tester l'homogénéité des $I_2 - 1$ modalités de la variable V_2 .

5.1.2.5 Modèle d'association "ligne-colonne"

Ici, on cherche à évaluer l'association à l'intérieur d'une cellule quelconque du tableau et non sur les différentes marges comme en témoignent les modèles (E 5.5) et (E 5.6). Deux aspects possibles de ce modèle sont concevables.

$$\theta_{i_2}(i_1, i_2) = \theta_{i_2}(+, i_2) \bullet \theta_{i_2}(i_1, +) \quad (\text{E 5.7a})$$

où

$$\begin{aligned} \text{Log} \theta_{i_2}(i_1, i_2) &= \theta_{i_2}(+, i_2) + \theta_{i_2}(i_1, +) & i_1=1, \dots, I_1-1 \\ & & i_2=1, \dots, I_2-1 \end{aligned} \quad (\text{E 5.7b})$$

Les deux modèles ont chacun $(I_1 - 1) + (I_2 - 1)$ degrés de contraintes soit $(I_1 - 2) (I_2 - 2)$ degrés de liberté. Tous les modèles (E 5.3), (E 5.4),... , (E 5.7a), (E 5.7b) correspondent aux modèles 0, U, R, C, RCI et RCII de Goodman (1979). Les cinq premiers sont des modèles Log-linéaires alors que le dernier ne l'est pas. En effet le modèle (E 5.7b) est Log-non linéaire. Ces différents modèles peuvent bien s'étendre à l'échelle multidimensionnelle. Nous avons limité leur présentation à l'échelle bidimensionnelle pour mieux cerner à ce stade, la similitude existant entre ces modèles et leur formulation Log-linéaire. On retrouvera ainsi chaque type de modèle défini ci-haut avec le même nombre de degrés de contraintes et par conséquent de degrés de liberté dans sa formulation Log-linéaire.

5.1.3. Formulation Log-linéaire des modèles d'association

Plusieurs cas de figure de ces modèles seront recensés ici, notamment leur formulation dans un tableau de dimension 2, dans un tableau de dimension 3 et l'extension dans un tableau de dimension p ($p > 3$).

5.1.3.1 Forme Log-linéaire des modèles d'association dans un tableau de dimension 2

Nous analyserons ici la forme que prend le modèle quand les variables du tableau sont nominale et ordinale ou toutes deux ordinales.

5.1.3.1 Modèle Log-linéaire appliqué à un tableau à variables nominale et ordinale

On rappelle encore que $t(i_1, i_2)$ est l'observation d'une cellule élémentaire et $t^*(i_1, i_2)$ la valeur espérée de $t(i_1, i_2)$ ($i_1 = 1, \dots, I_1$; $i_2 = 1, \dots, I_2$). D'autre part, les modèles d'indépendance et saturé sont les suivants et respectivement :

$$\text{Log} t^*(i_1, i_2) = a_0 + a_1(i_1) + a_2(i_2)$$

$$\text{Log} t^*(i_1, i_2) = a_0 + a_1(i_1) + a_2(i_2) + a_{i_2}(i_1, i_2) \quad (\text{E 5.8})$$

avec les contraintes ANOVA correspondantes (cf. E 4.11)

Les deux modèles de (E 5.8) font table rase sur l'information relative à l'ordre. Il nous faut donc développer un modèle qui en tienne compte. Pour cela, nous supposons que la variable V_1 (la première variable) du tableau est ordinale alors que la variable V_2 (la deuxième) est nominale. On assigne de façon arbitraire aux I_1 modalités de V_1 les scores S_i dans l'échelle de leur classement. Si celui-ci est monotone croissant alors on aura :

$$S_1 < S_2 < \dots < S_i < \dots < S_{I_1}$$

Soit \bar{S}_1 la moyenne arithmétique des I_1 scores

$$\bar{S}_1 = \frac{1}{I_1} \sum S_i \quad (\text{E 5.9})$$

L'opération de centrage n'altère pas la distance entre les scores (Cf. propriétés des modèles Log-non linéaires en (5.2.3.1.2). En transformant alors les scores par leur écart par rapport à la moyenne, on a comme modèle incorporant l'ordre des modalités de la variable V_1 un premier type de modèle tel que :

$$\begin{aligned} \text{Logit}^*(i_1, i_2) &= a_0 + a_1(i_1) + a_2(i_2) + (S_{i_1} - \bar{S}_1)a' \\ \sum a_1(i_1) &= \sum a_2(i_2) = 0 \end{aligned} \quad (\text{E 5.10})$$

a' est un terme constant .

Le terme $(S_{i_1} - \bar{S}_1)a'$ permet de mesurer l'association et pour une modalité quelconque de V_1 , l'association est la même quand on retient n'importe quelle modalité de V_2 . Ce modèle ressemble au modèle d'association uniforme sans pour autant l'être. Néanmoins, on peut dire qu'il est uniforme suivant la variable ordinale. $I_1 + I_2$ degrés de contraintes, soit $(I_1 - 1)(I_2 - 1) - 1$ degrés de liberté.

D'ailleurs, par rapport au modèle d'indépendance, un paramètre de plus est estimé. Il a donc autant de degrés de liberté que le modèle (E 5.4). Un autre modèle possible est tel que:

$$\begin{aligned} \text{Logit}^*(i_1, i_2) &= a_0 + a_1(i_1) + a_2(i_2) + (S_{i_1} - \bar{S}_1)a'_2(i_2) \\ \sum a_1(i_1) &= \sum a_2(i_2) = \sum a'_2(i_2) = 0 \end{aligned} \quad (\text{E 5.11})$$

$(S_{i_1} - \bar{S}_1)a'_2(i_2)$ mesure encore l'association et dépend de la variable ordinale à travers les scores S_{i_1} ($I_1=1, \dots, I_1$). Ce terme indique la déviation du modèle (E 5.11) par rapport au modèle d'indépendance (on retrouve l'idée qu'avait Jae-on Kim pour la mesure d'association, à la différence qu'ici, il n'y a pas de relation fonctionnelle sous-jacente). Les modèles (E 5.10) et (E 5.11) se situent entre le modèle d'indépendance et le modèle saturé.

Si $\alpha'_2(i_2) > 0$ (< 0), il y a croissance (décroissance) monotone de l'association à travers les I_1 modalités ordonnées de V_1 . Ainsi qu'on peut le noter dans la formulation proposée par Plackett (1981), $\alpha'_2(i_2)$ ($i_2 = 1, \dots, I_2$) peut être interprété comme la pente du terme linéaire décrivant l'association. Duncan (1979) interprète ces paramètres comme des scores des modalités de V_2 , estimés une fois que les modalités de V_1 ont été affectées des scores arbitraires.

Le modèle (E 5.11) n'est rien d'autre que le modèle (E 5.6.) traduit sous la forme Log-linéaire. Il a également $(I_1 - 2)(I_2 - 1)$ degrés de liberté, ce qui équivaut au nombre de degrés de liberté du modèle (E 5.6).

Si les modalités de V_2 plutôt que celles de V_1 étaient ordonnées, des scores S_{i_2} ($I_2=1, \dots, I_2$) tels que $S_1 < S_2 < \dots < S_{i_2} < \dots < S_{I_2}$ leur seraient assignés et on retrouverait le modèle ci-dessous :

$$\begin{aligned} \text{Logit}^*(i_1, i_2) &= a_0 + a_1(i_1) + a_2(i_2) + (S_{i_2} - \bar{S}_2)a'_2(i_2) \\ \sum a_1(i_1) &= \sum a_2(i_2) = \sum a'_2(i_2) = 0 \end{aligned} \quad (\text{E 5.12})$$

L'interprétation de $\alpha'_1(i_1)$ est symétriquement analogue à celle donnée à $\alpha'_2(i_2)$ et le modèle (E 5.12) est cette fois-ci similaire au modèle (E 5.5).

En appliquant le modèle (E 5.12) au tableau 2, on obtient les valeurs suivantes de $a'_1(i_1)$

$$\alpha'_1(1)=0.495 \quad \alpha'_2(2)=0.224 \quad \alpha'_3(3)=-0.719$$

On constate qu'il y a une échelle ordinale des valeurs obtenues de $\alpha'_1(i_1)$ et on peut conclure que les "Démocrates" forment le groupe le plus "libéral" par rapport aux deux autres, alors que les "républicains" apparaissent comme les plus "conservateurs".

Tableau 8 : Exemple d'estimation des modalités ordinales.

	Conservative	Moderate	Libéral	Total
Démocrate	100	156	143	399
Indépendant	140	210	119	470
Républicain	127	72	15	214
Total	368	438	277	1083

D'emblée, l'accent a été mis sur les modèles d'association se traduisant par un écart relatif par rapport à l'indépendance, ce qui ne doit pas occulter d'autres cas de figure où cet écart n'est pas apparent. Un modèle comme le suivant peut d'ailleurs être bâti :

$$\text{Logt}^*(i_1, i_2) = a_0 + a_1(i_1) + (S_{i_2} - \bar{S}_2)a' + (S_{i_2} - \bar{S}_2)a'_1(i_1) \quad (\text{E 5.13})$$

avec la variable V_2 supposée ordinale. Le terme $\alpha_2(i_2)$ n'apparaît plus dans le modèle et le modèle d'indépendance ne peut être un cas particulier de (E 5.13). Les paramètres α' et $\alpha'(i_1)$ s'interprètent toujours comme précédemment. Ce modèle n'a pas pour autant d'intérêt pratique dans la recherche des liaisons d'un tableau et il n'est surtout pas hiérarchique.

Remarques

Avec le modèle (E 5.12) par exemple, si $\alpha'_1(i_1) > 0$ (respectivement < 0) alors les effectifs en i_1 sont largement supérieurs (respectivement inférieurs) à ceux espérés en cas d'indépendance. En retenant deux modalités quelconques de V_1 (soit i_1^1 et i_1^2 pour une modalité fixée de V_2 , on a :

$$\text{Log}\theta_1(i_1, i_2) = (a_1(i_1^1) - a_1(i_1^2)) + (S_{i_2} - \bar{S}_2)(a'_1(i_1^1) - a'_1(i_1^2))$$

où $\theta_1(i_1, i_2)$ est le odd-ratio conditionnel de premier ordre.

On retrouve là encore une association linéaire entre i_1^1 et i_1^2 . Si $a'_1(i_1^1) > a'_1(i_1^2)$ alors l'association en i_1^1 est plus importante qu'en i_1^2 . D'autre part, en plus des deux modalités i_1^1 et i_1^2 , si l'on sélectionne deux autres modalités quelconques i_1^3 et i_1^4 de V_2 ($i_1^3 > i_1^4$), on forme

$$\text{Log}\theta_1(i_1, i_2) = (S_{i_2} - S_{i_2'}) (a'_1(i_1^1) - a'_1(i_1^2)) \quad (\text{E 5.15})$$

Alors que le modèle appliqué au odd-ratio conditionnel de premier ordre est de nature linéaire (cf. (E 5.14)), celui du odd-ratio de deuxième ordre est de nature multiplicative. Le facteur multiplicatif est l'écart - la distance- entre les scores. Si cet écart n'est pas connu, on retrouve le modèle Log-multiplicatif (cf. E 5.52)

Du modèle (E 5.15) la différence $(a'_1(i_1^1) - a'_1(i_1^2))$ mesure combien de fois on a de chance de se trouver en i_1^1 plutôt qu'en i_1^2 lorsque la variable V_2 est ordonnée. Dans l'exemple illustré plus haut avec le tableau 2 et auquel on applique le modèle (E 3.12), $\alpha'_1(1) - \alpha'_1(3) = 1.214$ c'est-à-dire qu'il y a $\exp(1,214) = 3,36$ fois plus de chance pour les "Démocrates" que pour les

"Républicains" d'être classés "Libéraux" par rapport aux "Modérés" plutôt que "Modérés " par rapport aux "Conservateurs".

L'analyse ne s'arrête pas au simple cas d'un tableau où l'une des variables est ordinale et l'autre nominale. Tout en restant toujours dans un tableau à deux dimensions, la forme du modèle lorsque toutes les deux variables du tableau sont ordinales sera elle aussi prospectée.

3.1.3.1.2 Modèle Log-linéaire appliqué à un tableau où toutes les deux variables sont ordinales

Dans le cas présent, les deux variables V_1 et V_2 du tableau sont toutes deux à modalités ordonnées. Deux sous ensembles de scores S_{i_2} ($I_2=1, \dots, I_2$) sont assignés aux modalités de V_1 et V_2 respectivement, suivant le schéma que nous avons vu pour les modèles (E 5.11) et (E 5.12). Un des modèles reflétant la forme ordonnée prise par les deux variables est le suivant :

$$\begin{aligned} \text{Logt}^*(i_1, i_2) &= a_0 + a_1(i_1) + a_2(i_2) + \gamma(S_{i_2} - \bar{S}_1)(S_{i_2} - \bar{S}_2) \\ \sum a_1(i_1) &= \sum a_2(i_2) = 0 \end{aligned} \quad (\text{E 5.16})$$

$\gamma(S_{i_2} - \bar{S}_1)(S_{i_2} - \bar{S}_2)$ est le terme d'association et comme dans (E 5.11) et (E 5.12), est une déviation par rapport à l'indépendance (lorsque $\gamma=0$, on retrouve le modèle d'indépendance). Si S_{i_1} et S_{i_2} sont strictement monotones croissants, l'association sera positive, négative ou nulle selon que γ est positif, négatif ou nul. Le modèle (E 5.16) peut se réduire à un modèle d'association uniforme (Cf. (E 5.4)), ce qui veut dire ici que l'association locale est la même à l'intérieur de chaque sous-tableau 2x2 formé à partir des modalités adjacentes de V_1 et V_2 (voir (E 5.18)). Chaque fois qu'une modalité quelconque de V_1 est fixée (respectivement V_2), le terme $\gamma(S_{i_2} - \bar{S}_1)(S_{i_2} - \bar{S}_2)$ est linéaire en V_2 (respectivement V_1). Le modèle est dit encore d'association linéaire-linéaire (HABERMAN (1974), AGRIESTI (1982), CLOGG(1982)). Ce modèle est évidemment à $(I_1-1)(I_2-1)$ degrés de liberté, soit encore autant que le modèle (E 5.4).

Si nous désignons par i_1^1 et i_1^2 deux modalités de V_1 et i_2^1 et i_2^2 deux modalités de V_2 , alors :

$$\begin{aligned} \text{Log} \theta_{i_2}(i_1, i_2) &= \gamma(S_{i_2^1} - S_{i_2^2})(S_{i_2^1} - S_{i_2^2}) \\ \text{avec : } i_1^1 &< i_1^2; i_2^1 < i_2^2 \end{aligned} \quad (\text{E 5.17})$$

Par conséquent, (E5.16) appliqué au odd-ratio du deuxième ordre est égal au produit de l'écart entre les scores affectés à V_1 et V_2 , au facteur multiplicatif près qui est γ .

GOODMAN (1979,1981) a étudié le cas simple où $s_{i_1} = i_1 (i_1 = 1, \dots, I_1)$ et $s_{i_2} = i_2 (i_2 = 1, \dots, I_2)$. Avec des modalités adjacentes, les écarts $(S_{i_1} - S_{i_1'})$ et $(S_{i_2} - S_{i_2'})$ sont tous identiques à l'unité (cas des modalités monotones croissantes et adjacentes). Ces conditions supplémentaires nous conduisent à la description de l'association uniforme :

$$\begin{aligned} \text{Log}\theta_{12}(i_1, i_2) &= \gamma \\ i_1 &= 1, \dots, I_1 - 1 \\ i_2 &= 1, \dots, I_2 - 1 \end{aligned} \quad (\text{E 5.18})$$

γ étant, bien sûr l'association uniforme (ici c'est le logarithme du odd-ratio par le produit d'unité de distance entre les modalités de V_1 et celles de V_2). Ce paramètre γ a une analogie non moins évidente avec le coefficient de régression dans la mesure où il indique le changement qui intervient lorsque l'on passe d'un sous tableau à l'autre quand les ordonnancements relatifs à V_1 et V_2 sont effectués. Néanmoins, malgré l'analogie, γ n'est pas le coefficient de régression au sens strict.

Il est possible d'utiliser la combinaison des modèles (E 5.11) et (E5.12) pour le traitement d'un tableau de dimension 2, d'où :

$$\text{Logt}^*(i_1, i_2) = a_0 + a_1(i_1) + a_2(i_2) + (s_{i_2} - \bar{s}_2)a'_1(i_1) + (s_{i_1} - \bar{s}_1)a'_2(i_2)$$

$$\sum a_2 i(2) = \sum a_1 i(1) = \sum a'_2 i(2) = \sum a'_1 i(1) = 0 \quad (\text{E 5.19})$$

Ce modèle permet de mesurer simultanément les associations (elles sont linéaires) relatives à V_1 (c'est le terme le terme $(s_{i_2} - \bar{s}_2)a'_1(i_1)$ et V_2 (le terme $(s_{i_1} - \bar{s}_1)a'_2(i_2)$). Il a $(I_1-2)(I_2-2)-1$ degrés de liberté. La somme des deux associations indique également la déviation par rapport au modèle d'indépendance. Un autre modèle appliqué au tableau à variables ordinales et qui ne dévie pas de l'indépendance serait:

$$\text{Logt}^*(i_1, i_2) = a_0 + a_1(i_1) + a_2(i_2) + (s_{i_2} - \bar{s}_2)a' + \gamma(s_{i_1} - \bar{s}_1)(s_{i_2} - \bar{s}_2) \quad (\text{E 5.20})$$

Le terme $\alpha_2(i_2)$ n'apparaît pas dans le modèle (E 5.20), ce qui signifie également que le modèle d'indépendance ne peut être un cas particulier de (E 5.20). Ce dernier modèle a $I_1 + 2$ degrés de contraintes.

Toutes les analyses conduites ici dans le tableau de dimension 2 peuvent être transposées dans un tableau de dimension 3, voire de dimension supérieure à 3, juste en procédant au même raisonnement que dans un tableau de dimension 2.

5.1.3.2 Forme Log-linéaire des modèles d'association dans un tableau de dimension 3

En présence d'un tableau de dimension 3, constitué des variables V_1 , V_2 et V_3 à I_1 , I_2 et I_3 modalités respectives, trois cas de figures faisant intervenir des variables ordinales peuvent se Présenter :

- cas d'un tableau à une variable ordinale et les deux autres nominales.
- cas d'un tableau à deux variables ordinales et la troisième nominale,
- cas d'un tableau où toutes les trois variables sont ordinales.

Le cas de figure du tableau où toutes les trois variables sont nominales est écarté ici (il ne fait pas l'objet de l'étude dans ce paragraphe) puisqu'il est déjà décrit par la forme Log-linéaire traditionnelle ne prenant pas en compte l'ordre que reflètent les modalités des variables (cf. chap. 4, E 4.12)

L'observation $t^*(i_1, i_2, i_3)$ ($i_1 = 1, \dots, I_1$; $i_2 = 1, \dots, I_2$; $i_3 = 1, \dots, I_3$) est la valeur espérée de l'observation $t(i_1, i_2, i_3)$ à l'intérieur d'une cellule élémentaire du tableau de dimension 3. Les différents cas de figure esquissés seront successivement examinés ici.

5.1.3.2.1. Modèle Log-linéaire appliqué à un tableau à une variable ordinale et deux variables nominales

Supposons que la variable V_3 est ordinale. En opérant de la même façon qu'en 5.1.3.1, on assigne arbitrairement des scores s_{i_3} ($i_3 = 1, \dots, I_3$) aux modalités de V_3 . S'ils sont monotones croissants, c'est-à-dire tels que $s_1 < s_2 < \dots < s_{I_3}$, alors un premier type de modèle proposé est décrit par :

$$\text{Logit}^*(i_1, i_2, i_3) = a_0 + a_1(i_1) + a_2(i_2) + a_3(i_3) + (s_{i_3} - \bar{s}_3)a'_1(i_1) + (s_{i_3} - \bar{s}_3)a'_2(i_2) \quad (\text{E 5.21})$$

avec les contraintes appropriées. Ce modèle traduit l'existence d'une indépendance partielle entre V_1 et V_2 de même que l'association, mesurée par les termes $(s_{i_3} - \bar{s}_3)a'_1(i_1)$ et $(s_{i_3} - \bar{s}_3)a'_2(i_2)$ est, une nouvelle fois, une déviation par rapport à l'indépendance totale. A l'inverse, s'il y avait dépendance partielle entre V_1 et V_2 , le modèle (E5.21), avec l'inclusion d'un terme supplémentaire $a_{12}(i_1, i_2)$ deviendrait le suivant :

$$\begin{aligned} \text{Logit}^*(i_1, i_2, i_3) &= a_0 + a_1(i_1) + a_2(i_2) + a_3(i_3) + a_{12}(i_1, i_2) + (s_{i_3} - \bar{s}_3)a'_1(i_1) + (s_{i_3} - \bar{s}_3)a'_2(i_2) \\ \sum a_2 i(2) &= \sum a_1 i(1) = \sum a'_2 i(2) = \sum a'_1 i(1) = 0 \\ \sum_{i_1} a_{21}(i_1, i_2) &= \sum_{i_2} a_{12}(i_1, i_2) = 0 \end{aligned} \quad (\text{E 5.22a})$$

Le modèle (E 5.22a), contrairement au modèle (E5.21) n'est pas une déviation par rapport au modèle d'indépendance dans le sens défini haut. Il est intéressant de souligner ici que les termes de mesure d'association respectent le principe des modèles hiérarchiques (un terme d'effet d'ordre supérieur ne peut exister que s'il existe des termes d'ordre inférieur l'impliquant). Prenons à titre d'exemple le terme $(s_{i_3} - \bar{s}_3)a'_1(i_1)$ et supposons qu'il y a dépendance partielle entre V_1 et V_2 . On peut donc trouver un terme intermédiaire entre $a_{12}(i_1, i_2)$ et $a_{123}(i_1, i_2, i_3)$ qui respecte le principe hiérarchique et la notion d'ordre prise par les modalités de V_3 . Du fait de l'inclusion dans le modèle de $(s_{i_3} - \bar{s}_3)a'_1(i_1)$, le terme intermédiaire sera ici $(s_{i_3} - \bar{s}_3)a'_{12}(i_1, i_2)$. En partant de ce principe, plusieurs types de modèles peuvent être constitués de façon à permettre la prise en compte de l'ordre lié aux modalités de V_3 et chaque fois seule l'application des techniques de choix de modèles pourra permettre d'en retenir le "meilleur". Nous n'en sommes pas là pour l'instant. Toutefois, en appliquant le principe hiérarchique au modèle (E 5.22), celui-ci se transforme et on le réécrit suivant :

$$\begin{aligned} \text{Logit}^*(i_1, i_2, i_3) &= a_0 + a_1(i_1) + a_2(i_2) + a_3(i_3) + a_{12}(i_1, i_2) + (s_{i_3} - \bar{s}_3)a'_1(i_1) + (s_{i_3} - \bar{s}_3)a'_2(i_2) \\ &\quad + (s_{i_3} - \bar{s}_3)a'_{12}(i_1, i_2) \end{aligned}$$

avec les contraintes appropriées.

En quittant le cas de figure qui vient d'être abordé, il devient normal de présenter celui dans lequel se pose l'étude d'un tableau à deux variables ordinales et une nominale

5.1.3.2.2 Modèle Log-linéaire appliqué à un tableau à deux variables ordinales et une variable nominale

Soient V_2 et V_3 les variables ordinales et auxquelles on assigne des scores arbitraires

s_{i_2} ($i_2 = 1, \dots, I_2$) ; s_{i_3} ($i_3 = 1, \dots, I_3$). Un des modèles concevables a la structure ci-après :

$$\begin{aligned} \text{Logit}^*(i_1, i_2, i_3) = & a_0 + a_1(i_1) + a_2(i_2) + a_3(i_3) + (s_{i_2} - \bar{s}_2)a_1^1(i_1) + \phi(s_{i_3} - \bar{s}_3)a_1^1 + (s_{i_3} - \bar{s}_3)a_1^2(i_1) \\ & + \phi(s_{i_3} - \bar{s}_3)a_1^2(i_1) + \phi(s_{i_1} - \bar{s}_1)(s_{i_2} - \bar{s}_2) \end{aligned} \quad (\text{E 5.23})$$

avec les contraintes ANOVA appropriées.

$a_1^1(i_1)$ est le paramètre de mesure d'association relative à V_1 quand la variable V_2 est ordonnée.

$a_1^2(i_1)$ est le paramètre de mesure d'association relative à V_1 quand la variable V_3 est ordonnée

ϕ mesure l'association conditionnelle entre V_2 et V_3 .

Les trois derniers termes de (E 5.23) mesurent chacun l'association partielle et on constate encore ici que (E 5.23) dévie de l'indépendance totale.

Dans ce modèle, comme les précédents, toutes les contraintes de marges ne sont pas respectées (on reviendra ultérieurement au problème d'estimation de ces modèles pour ajouter d'autres contraintes supplémentaires de marges). Du modèle (E 5.23) par exemple, seules les marges de premier ordre (contraintes internes) sont respectées suivant le principe de BIRCH. Pour que la contrainte $t(i_1, i_2, +) = t^*(i_1, i_2, +)$ soit vérifiée, il faudrait dans ce cas qu'intervienne dans le modèle (E 5.23) le terme d'effet de paramètre $a_{12}(i_1, i_2)$, ce qui permet d'obtenir le modèle suivant :

$$\begin{aligned} \text{Logit}^*(i_1, i_2, i_3) = & a_0 + a_1(i_1) + a_2(i_2) + a_3(i_3) + a_{12}(i_1, i_2) + (s_{i_2} - \bar{s}_2)a_1^1(i_1) + \\ & (s_{i_3} - \bar{s}_3)a_1^2(i_1) + \phi(s_{i_2} - \bar{s}_2)(s_{i_3} - \bar{s}_3) \end{aligned}$$

$$\begin{aligned} \sum a_2(i_1) = \sum a_2(i_2) = \sum a_3(i_3) = \sum a_1^1(i_2) = \sum a_1^2(i_1) = 0 \\ \sum_{i_1} a_{12}(i_1, i_2) = \sum_{i_2} a_{12}(i_1, i_2) = 0 \end{aligned} \quad (\text{E5.24})$$

Si $s_{i_2} = i_2$ et si $s_{i_3} = i_3$, on peut former pour tout couple (i_2^1, i_2^2) de V_2 et (i_3^1, i_3^2) de V_3 $((i_2^1 > i_2^2); (i_3^1 > i_3^2))$ le odd-ratio conditionnel d'ordre 2 qui est $\theta_{123}(i_1, i_2, i_3)$ (odd-ratio conditionnel des variables V_2 et V_3 à travers les modalités de V_1 (nominale)).

Les modèles (E 5.23) ou (E 5.24), appliqués à cet odd-ratio donnent :

$$\text{Log} \theta_{123}(i_1, i_2, i_3) = \phi(s_{i_2^2} - s_{i_2^1})(s_{i_3^2} - s_{i_3^1}) \quad (\text{E 5.25})$$

pour des valeurs adjacentes i_2^2 et i_2^1 d'une part, i_3^1 et i_3^2 , d'autre part. On se trouve alors dans la situation du modèle (E 5.18) où il est question d'association uniforme.

V_1 et V_2 ont été supposées ordinales. Si c'était le cas pour V_1 et V_2 seulement, au lieu de (E5.23), on obtiendrait le modèle :

$$\begin{aligned} \text{Log} t^*(i_1, i_2, i_3) = & a_0 + a_1(i_1) + a_2(i_2) + a_3(i_3) + (s_{i_1} - \bar{s}_1)a_3^1(i_3) + \\ & (s_{i_2} - \bar{s}_2)a_2^3(i_3) + \phi'(s_{i_1} - \bar{s}_1)(s_{i_2} - \bar{s}_2) \end{aligned} \quad (\text{E 5.26})$$

avec bien entendu, toutes les contraintes nécessaires sur les paramètres. Il existe encore plusieurs combinaisons possibles de modèles et on ne saurait clore ce paragraphe sans expliciter la forme du modèle lorsque toutes les trois variables sont ordinales

5.1.3.2.3 Modèle Log-linéaire appliqué à un tableau ayant trois variables ordinales

La situation présente se caractérise par l'ordinalité simultanée de V_1 , V_2 et V_3 . L'assignation des scores, s_{i_2} et s_{i_3} à ces variables permet de formuler le modèle (E 5.27) tel que :

$$\text{Log} t^*(i, i, i) = a_0 + a_1(i) + a_2(i) + a_3(i) + \gamma_1(S_{i_1} - \bar{S}_1)(S_{i_2} - \bar{S}_2) + \gamma_2(S_{i_1} - \bar{S}_1)(S_{i_3} - \bar{S}_3) + \gamma_3(S_{i_2} - \bar{S}_2)(S_{i_3} - \bar{S}_3) \quad (\text{E 5.27})$$

La déviation de (E 5.27) de l'indépendance reste une fois encore apparente

($\gamma_1 = 0$; $\gamma_2 = 0$; $\gamma_3 = 0$, ce qui implique l'indépendance). Ce modèle a $(I_1-1)(I_2-1)(I_3-1)-3$ degrés de liberté (soit trois paramètres supplémentaires à estimer par rapport au modèle d'indépendance). Un autre modèle, moins parcimonieux que (E5.27) prendra la forme suivante :

$$\begin{aligned} \text{Logit}^*(i, i, i) = & a_0 + a_1(i) + a_2(i) + \alpha_3(i) + \gamma_1(S_{i.} - \bar{S}_{i.})(S_{.i} - \bar{S}_{.i}) + \gamma_2(S_{i.} - \bar{S}_{i.})(S_{.i} - \bar{S}_{.i}) + \gamma_3(S_{i.} - \bar{S}_{i.})(S_{.i} - \bar{S}_{.i}) \\ & \gamma_4(S_{i.} - \bar{S}_{i.})(S_{.i} - \bar{S}_{.i})(s_i - \bar{s}_i) \end{aligned} \quad (\text{E 5.28})$$

Contrairement à ce que peut évoquer cette formulation, le modèle (E 5.28) n'est pas saturé et se trouve toujours intermédiaire entre le modèle d'indépendance totale et le modèle saturé. D'ailleurs, celui-ci a $I_1 * I_2 * I_3$ degrés de contraintes alors que (E 5.23) n'en a que $I_1 + I_2 + I_3 + 2$. L'exclusion des différentes associations partielles de (E 5.28) se traduit par :

$$\text{Logit}^*(i, i, i) = a_0 + a_1(i) + a_2(i) + \alpha_3(i) + \gamma_4(S_{i.} - \bar{S}_{i.})(S_{.i} - \bar{S}_{.i})(s_i - \bar{s}_i) \quad (\text{E 5.29})$$

Le modèle (E 5.29) permet de mesurer l'association en même temps qu'il incarne un type précis de modèle : le modèle non hiérarchique, qui ne nous préoccupe pas pour autant dans la présente étude et qui aurait mérité une étude assez approfondie.

Pour des modalités adjacentes V_2 et de V_3 , si l'on forme les différents odd-ratios conditionnels et leur applique le modèle (E5.27), on a finalement :

$$\begin{aligned} \text{Log } \theta_{1,2}(i, i, i) &= \gamma_1(s_i - s_{i'}) (s_i - s_{i'}) \\ \text{Log } \theta_{1,3}(i, i, i) &= \gamma_2(s_i - s_{i'}) (s_i - s_{i'}) \\ \text{Log } \theta_{2,3}(i, i, i) &= \gamma_3(s_i - s_{i'}) (s_i - s_{i'}) \end{aligned} \quad (\text{E 5.30a})$$

De plus, lorsque $s_{i.} = i$, $s_{.i} = i$, $s_{i.} = i$ et $i' > i^*$, $i' > i^*$, $i' > i^*$ alors ces équations se ramènent à :

$$\begin{aligned} \text{Log } \theta_{1,2}(i, i, i) &= \gamma_1 \\ \text{Log } \theta_{1,3}(i, i, i) &= \gamma_2 \\ \text{Log } \theta_{2,3}(i, i, i) &= \gamma_3 \end{aligned} \quad (\text{E 5.30b})$$

Les équations (E5.30b) ont bien évidemment été déjà rencontrées précédemment. Il s'agit des mesures d'association partielle uniforme.

Du mode le (E 5.28), γ_4 représente l'interaction et non l'association. Compte tenu des hypothèses requises pour aboutir aux équations (E530b), l'expression du odd-ratio du troisième ordre sous (E 5.28) se définit par :

$$\text{Log } \theta_{123}(i, i, i) = \gamma_4 \quad (\text{E 5.31})$$

γ_4 caractérisant l'interaction uniforme : on dira par exemple que l'association entre V_1 et V_2 est identique quel que soit le niveau (la modalité) de V_3 . Du fait de la symétrie, la conclusion est la même pour l'association entre V_1 et V_3 suivant les niveaux de V_2 ou l'association entre V_2 et V_3 suivant les niveaux de V_1 . Il y a néanmoins des situations où l'association entre deux variables peut être hétérogène à travers les modalités d'une troisième variable. On s'en aperçoit par l'application de (E 5.27) au odd-ratio conditionnel.

$$\text{Log } \theta_{123}(i_1, i_2, i_3) = \gamma_4 (s_{i_1} - \bar{s}_3) \quad (\text{E 5.32})$$

(E 5.32) est un modèle dans lequel l'association entre V_1 et V_2 est hétérogène t (n'est pas identique) à travers toutes les modalités de V_3 .

Toutes les analyses que nous venons de voir dans ce paragraphe s'étendent facilement aux tableaux de dimension p ($p > 3$) et constitués à la fois de variables ordinales et nominales ou ordinales uniquement. Chaque fois, on pourra toujours trouver un modèle assez parcimonieux, se situant entre le modèle d'indépendance et le modèle saturé et reflétant la nature ordonnée des variables. Les termes d'association seront interprétés de la même manière que dans les modèles (E 5.8) (E 5.9), ... , (E 5.32).

Il existe différents autres types de modèles servant au traitement des tableaux ayant des variables ordinales. C'est le cas du modèle logit, du modèle fondé sur les polynômes orthogonaux (Haberman, (1974)) et du modèle Log-non linéaire, lesquels seront présentés dans le paragraphe que nous ouvrons actuellement.

5.2. AUTRES TYPES DE MODELES COMME OUTILS DE TRAITEMENT DES TABLEAUX A VARIABLES ORDINALES

Outre l'approche Log-linéaire développée en 5.1, plusieurs autres formes de modèles peuvent être indispensables pour le traitement des tableaux de contingence à variables

ordinales. Tous ces modèles ne seront pas abordés ici et nous nous limiterons particulièrement à la présentation de trois d'entre eux, du moins dans leur forme prenant compte la caractéristique ordre.

- l'approche logit (cf. 5.2.1) du traitement des tableaux à variables ordinales et qui apparaît comme un cas particulier de l'approche Log-linéaire (cf. 5.1). D'ailleurs, il existe une parfaite similitude quant aux degrés de liberté fournis par certains modèles dans les deux approches. L'une des caractéristiques de l'approche logit est qu'elle est encore plus parcimonieuse que l'approche Log-linéaire.

- la paramétrisation des tableaux à variables ordinales par les doubles polynômes orthogonaux (5.2.2). Il ressortira de ce modèle que la régression par double polynômes orthogonaux peut être considérée comme un modèle Log-linéaire. Les paramètres estimés par cette paramétrisation seront caractérisés par leur unicité.

- le modèle Log-non linéaire (5.2.3) qui est la forme Log-linéaire des modèles d'association lorsque les scores associés aux modalités des variables ordinales ne sont pas fixés d'avance.

5.2.1 Forme Logit des modèles d'association

Le modèle logit est un modèle qui s'applique généralement à un tableau dans lequel l'une des variables est supposée dépendante et l'autre (les autres) indépendante (s) (cas d'un tableau de dimension supérieure ou égale à 2). D'autre part, ce modèle est couramment utilisé lorsque la variable dépendante est dichotomique. Avant de le présenter, il est clair de noter, afin de lever toute ambiguïté probable, que le modèle logit met en relief une relation causale entre la variable que nous supposons dépendante et celles supposées indépendantes, mais qu'il n'est pas de même nature que le modèle de régression simple (cf. 3.5.1)

Dans un tableau à deux dimensions, si l'on suppose la deuxième variable dichotomique et dépendante, le modèle logit se traduit par :

$$\log \frac{r^{*(i,1)}}{r^{*(i,2)}} = a_0 + a(i) \quad (\text{E 5.33})$$

$$\text{avec : } \sum a(a) = 0$$

Bien que spécifiant une structure nouvelle du tableau à deux dimensions, le modèle (E5.33) est équivalent au modèle saturé dans un tableau bidimensionnel (cf. (E 4.10)).

Le terme (a_i) du modèle (E 5.33) exprime l'interaction, ce qui n'est pas le cas dans le modèle Log-linéaire et si $a_i = 0 \forall i$, alors il y a indépendance. Sous cette forme, le modèle (E 3.33) ne tient pas compte de l'ordre si l'une des variables est ordinale. De plus, quoiqu'il soit couramment utilisé pour les variables dichotomiques, il peut être aussi appliqué aux variables polytomiques (Fienberg, 1980), (O.D. Williams et JE, Grizzle, 1972). Il existe plusieurs formes de modèles logits appliquées aux variables polytomiques dont nous retiendrons celle décrite par Agresti (1982), qui est le modèle logit "accumulé". Ainsi, dans un tableau à deux dimensions, la fixation de la variable V_1 comme dépendante et la variable V_2 comme indépendante entraîne la construction de $I_1 - 1$ logits accumulés (V_1 à I_1 modalités ; V_2 en I_2). On le verra plus explicitement dans les paragraphes qui suivent, ou nous examinerons la forme prise par ce modèle dans des tableaux de dimension 2 et 3.

5.2.1.1 Formulation logit du modèle d'association dans un tableau de dimension 2

Le cadre d'étude reste un tableau à deux dimensions, où toutes les variables sont polytomiques, ce qui exclue du champ d'analyse la forme dichotomique généralement utilisée. En supposant $t^*(i_1, i_2)$ la valeur estimée de $t(i_1, i_2)$ par un quelconque modèle, le $i_1^{\text{ème}}$ logit « accumulé » ($i_1 = 1, \dots, I_1$) s'exprimera par :

$$\text{Log} \left[\frac{t^*(i+1, i) + t^*(i+2, i) + \dots + t^*(I, i)}{t^*(1, i) + t^*(2, i) + \dots + t^*(i, i)} \right] = \text{Log} \frac{\sum_{i=1}^{I-i} t^*(1, i)}{\sum_{i=1}^I t^*(1, i)} \quad \begin{array}{l} i=1, \dots, I-1 \\ i=1, \dots, I \end{array} \quad (\text{E 5.34})$$

Posons pour la suite :

$$L(i, i) = \text{Log} \text{Log} \frac{\sum_{i=1}^{I-i} t^*(1, i)}{\sum_{i=1}^I t^*(1, i)}$$

$L(i_1, i_2)$ étant comme on peut le constater le $i_1^{\text{ème}}$ logit « accumulé ». Comme dans l'approche Log-linéaire, deux cas de figure se conçoivent :

- cas d'un tableau ayant une variable ordinale et une variable nominale,
- cas d'un tableau où toutes les deux variables sont ordinales.

5.2.1.1.1 Cas d'un tableau à une variable ordinale et une variable nominale

Supposons que la variable V_1 est ordinale et considérée dépendante. De même que dans le modèle Log-linéaire, des scores seront assignés aux modalités de V_1 pour décrire la

structure du modèle. Mais ici, cette assignation de scores reste inutile, car la variable V_1 , ordinaire est la variable de réponse. La forme simple que prendra le modèle sera :

$$\begin{aligned} L(i, i_1) &= u_i + a_i(i_1) \\ \sum a_i(i_1) &= 0 \quad i_1 = 1, \dots, I_1 - 1 \\ & \quad i_1 = 1, \dots, I_1 \end{aligned} \quad (E 5.35a)$$

Ce qui nous ramène presque à (E 5.33), mais avec une différence quant au terme u_i (par rapport à α_0) qui ici est une moyenne arithmétique des différents Logits "accumulés", pour un niveau i_1 donné et à travers les I_2 modalités de V_2 c'est-à-dire que :

$$u_i = \frac{\sum L(i, i_1)}{I_2} \quad i_1 = 1, \dots, I_1 - 1 \quad (E 5.35b)$$

$\alpha_2(i_2)$ mesure également l'association.

Le modèle (E 5.35a) possède $I_1 - I_2 - 2$ degrés de contraintes. Il a donc $(I_1 - 2) (I_2 - 1)$ degrés de liberté, soit autant que le modèles Log-linéaire (E 5.11), bien que les degrés de contraintes des deux modèles soient différents.

On peut former le odd-ratio d'ordre 2 « accumulé » pour deux modalités quelconques i_1 et i_2 de V_2 :

$$\theta_{i_1, i_2} = \left[\frac{t^*(i_1 + 1, i_2) + \dots + t^*(I_1, i_2)}{t^*(1, i_2) + \dots + t^*(i_1, i_2)} \right] / \left[\frac{t^*(i_1 + 1, i_1) + \dots + t^*(I_1, i_1)}{t^*(1, i_1) + \dots + t^*(i_1, i_1)} \right] \quad (E 5.36)$$

Ce odd-ratio, lissé par le modèle (E 5.35a) donne :

$$\begin{aligned} \text{Log } \theta_{i_1, i_2} &= \text{Log } \frac{t^*(i_1 + 1, i_2) + \dots + t^*(I_1, i_2)}{t^*(1, i_2) + \dots + t^*(i_1, i_2)} - \text{Log } \frac{t^*(i_1 + 1, i_1) + \dots + t^*(I_1, i_1)}{t^*(1, i_1) + \dots + t^*(i_1, i_1)} \\ &= L(i_1, i_2) - L(i_1, i_1) \\ &= a_i(i_2) - a_i(i_1) \end{aligned}$$

Par rapport à (E 5.15) il n'y a plus de facteur multiplicatif ici (les scores dans le cas (E 5.15)), puisque V_1 ordonnée est en (E 5.35a) une variable de réponse.

Lorsque $a_i(i_2) > a_i(i_1)$, l'association à la marge $M(i_2)$ (ou la distribution conditionnelle en i_1) est plus importante qu'en i_1

Le modèle (E 5.35a) se situe toujours entre le modèle d'indépendance (cas où $a_i(i) = 0 ; i=1, \dots, I_1$) et le modèle saturé dont la structure est telle que :

$$L(i, i) = u_i + a_i(i, i) \quad (E 5.37)$$

Le modèle d'indépendance qui découle donc du modèle (E 5.35a) lorsque $a_i(i) = 0 \forall i_2$, se ramène à un modèle permettant de tester l'homogénéité des modalités de V_2 . On retrouve ici l'analyse faite par Gary Simon (1974) dans sa formulation B.

La formulation du modèle logit se conçoit aussi dans un tableau ayant toutes les deux variables ordinales.

5.2-1.1-2. Cas d'un tableau à deux variables ordinales

On est donc en présence d'un tableau de dimension 2 où toutes les deux variables sont ordinales et comme il a été expliqué tout à l'heure, les scores seront affectés cette fois-ci seulement aux modalités de V_2 et non à celles de V_1 qui est toujours la variable de réponse. Ces scores étant supposés monotones croissants ($s_1 < s_2 < \dots < s_n$), il existe encore $(I_1 - 1)$ modèles logits « accumulés » dont la décomposition est

$$L(i, i) = u_i + \gamma(s_n - \bar{s}_i) \quad \begin{matrix} i=1, \dots, I_1 - 1 \\ i=1, \dots, I_2 \end{matrix} \quad (E 5.38)$$

Chaque logit est lié linéairement à la variable V_2 avec une pente constante identique pour tous les $(I_1 - 1)$ logits "accumulés. De (E 5.38) seule le terme u_i varie d'un logit à l'autre. Ce terme est calculé de la même manière qu'en (E 5.35a) et l'indépendance est un cas particulier de (E 5.38) puisque le modèle d'indépendance s'obtient pour $\gamma = 0$ (ce n'est plus le cas d'homogénéité de tout à l'heure).

Par le modèle (E5.35a), si la variable V_1 , ordinaire n'était pas prise pour variable de réponse, le modèle prendrait la forme qui suit :

$$L(i, i) = u_i + \gamma(s_n - \bar{s}_i) \quad \begin{matrix} i=1, \dots, I_1 - 1 \\ i=1, \dots, I_2 \end{matrix} \quad (E 5.39)$$

$$u_i = \frac{\sum L(i, i)}{I_2}$$

$s_n (i=1, \dots, I_1)$ étant les scores affectés aux modalités de V_1 . A l'inverse, dans le modèle (E 5.38), si la variable V_2 est la variable de réponse, on retrouve encore le modèle (E 5.39). Il semble

donc, qu'on arrive au même modèle, chaque fois que la variable ordinale n'est pas la variable de réponse (cas d'un tableau à une variable nominale et une variable ordinale ou bien le cas d'un tableau à deux variables ordinales), le modèle (E5.38) illustrant une telle situation. Revenons au modèle (E5.38). Soient deux modalités fixées i_1^j et i_2^j de V_2 avec lesquelles on construit le odd-ratio "accumulé" d'ordre 2. (Voir (E5.36a)).

Le modèle (E 5.38) appliqué à ce odd-ratio est tel que :

$$\text{Log } \theta_2(i_1, i_2) = L(i_1, i_2^j) - L(i_1, i_2^k) = \gamma(s_{i_1} - s_{i_2}) \quad (\text{E 5.40a})$$

Si les scores s_{i_1} et s_{i_2} sont consécutifs et tels que $s_{i_1} - s_{i_2} = 1$, alors

$$\text{Log } \theta_2(i_1, i_2) = \gamma \quad (\text{E 5.40b})$$

γ étant la mesure de l'association uniforme.

Le modèle (E5.38) a I_1 degrés de contraintes, c'est-à-dire $(I_1 - 1)(I_2 - 1) - 1$ degrés de liberté, soit autant que le modèle (E 5.16). De la même manière qu'avec le modèle Log-linéaire, l'extension du modèle logit à l'étude des tableaux multidimensionnels ayant des variables ordinales est possible et dans le paragraphe suivant, nous l'appliquerons plus particulièrement à un tableau de dimension 3.

5.2.1.2. Formulation logit du modèle d'association dans un tableau de dimension 3

Ici, nous retiendrons comme en 5.1.3.2, trois cas de figure, c'est-à-dire les cas d'un tableau où l'on trouve :

- une variable ordinale et deux variables nominales,
- deux variables ordinales et une variable nominale,
- toutes les trois variables ordinales.

Chacun des cas sera rapidement décrit, ceci pour la simple raison que le modèle logit n'a pas autant de termes dans sa décomposition que le modèle log-linéaire, ce qui aurait permis d'entreprendre différentes formes d'analyse du modèle dans chaque cas.

Envisageons le premier cas de figure où il n'y a qu'une variable ordinale dans le tableau. Soit V_1 cette variable et supposons qu'elle est également la variable de réponse.

Le $i_1^{\text{ème}}$ logit "accumulé" noté $L(i_1, i_2, i_3)$ se définit par:

$$L(i_1, i_2, i_3) = \text{Log} \left[\frac{t^*(i_1+1, i_2, i_3) + \dots + t^*(I_1, i_2, i_3)}{t^*(1, i_2, i_3) + \dots + t^*(i_1, i_2, i_3)} \right]$$

$$i_1 = 1, \dots, I_1 - 1$$

$$i_2 = 1, \dots, I_2 \quad (\text{E 5.41a})$$

$$i_3 = 1, \dots, I_3$$

$L(i_1, i_2, i_3)$ a la décomposition suivante :

$$L(i_1, i_2, i_3) = u_{i_1} + a_2(i_2) + a_3(i_3)$$

$$\sum a_2(i_2) = \sum a_3(i_3) = 0$$

$$u_{i_1} = \frac{\sum L(i_1, i_2, i_3)}{I_2 I_3} \quad i_1 = 1, \dots, I_1 - 1 \quad (\text{E 5.41b})$$

$\alpha_2(i_2)$ et $\alpha_3(i_3)$ étant des effets mesurant l'association partielle) entre V_1 et V_2 d'une part, et V_1 et V_3 d'autre part. Le modèle (E 5.41b) fait abstraction de l'existence d'une liaison de dépendance conditionnelle entre V_2 et V_3 . Si cette liaison existait, il faudrait inclure au modèle (E 5.41b) le terme $\alpha_{23}(i_2, i_3)$ et dans ce cas, on obtiendrait :

$$L(i_1, i_2, i_3) = u_{i_1} + a_2(i_2) + a_3(i_3) + a_{23}(i_2, i_3)$$

$$\sum a_2(i_2) = \sum a_3(i_3) = \sum a_{23}(i_2, i_3) = 0 \quad (\text{E 5.42})$$

- Si V_1 et V_2 sont ordinales, V_1 jouant toujours le rôle de la variable de réponse. alors en assignant les scores aux modalités de V_2 et en définissant $L(i_1, i_2, i_3)$ toujours par:

$$L(I_1, I_2, I_3) = \text{Log} L(i_1, i_2, i_3) = \text{Log} \left[\frac{\sum_{i_2 \geq i_1} t^*(1, i_2, i_3)}{\sum_{i_2 \leq i_1} t^*(1, i_2, i_3)} \right]$$

On a un modèle dont la forme structurelle est telle que

$$L(i_1, i_2, i_3) = u_{i_1} + a_2(i_2) + a_3(i_3) + \gamma_2(s_{i_2} - \bar{s}_2) \quad (\text{E 5.43})$$

$\gamma_2(s_{i_2} - \bar{s}_2)$ indique l'association linéaire entre V_1 et V_2 avec une pente constante (γ_2 mesure d'association uniforme dans le cas des modalités adjacentes) pour tous les modèles logits.

Lorsque le tableau possède toutes les trois variables ordinales, l'assignation des scores s_{i_2} ($i_2 = 1, \dots, I_2$) à la variable V_2 et des scores s_{i_3} ($i_3 = 1, \dots, I_3$) à la variable V_3 transforme le

terme $a_3(i_3)$ (mesure d'association entre V_1 et V_3) du modèle (E 5.43) en $\gamma_3(s_{i_3} - \bar{s}_3)$ et le nouveau modèle se présente suivant :

$$L(i_1, i_2, i_3) = u_{i_1} + \gamma_2(s_{i_2} - \bar{s}_2) + \gamma_3(s_{i_3} - \bar{s}_3)$$

$$i_1=1, \dots, I_1-1$$

$$i_2=1, \dots, I_2 \quad (E 5.44)$$

$$i_3=1, \dots, I_3$$

Le terme $\gamma_3(s_{i_3} - \bar{s}_3)$ mesure l'association linéaire entre V_1 et V_2 avec une pente constante γ_3 pour les modèles logits, ceteris paribus.

On constate à la lumière de ces modèles que l'approche logit des modèles d'association conduit à des modèles plus parcimonieux que dans l'approche Log-linéaire. Au-delà, il existe également une autre approche des modèles d'association (Log-linéaire) fondée sur les doubles polynômes orthogonaux.

5.2.2 Une paramétrisation du modèle Log-linéaire fondée sur les polynômes orthogonaux

HABERMAN (1974) est l'un de ceux qui ont développé l'approche du modèle Log-linéaire fondée sur les polynômes orthogonaux et qui est fort bien utilisée pour l'étude de la structuration d'un tableau à variables ordinales. Cette approche, bien que peu différente de celle présentée au paragraphe (5.1.3) - on assigne toujours des scores aux modalités ordonnées des variables - s'en distingue néanmoins car elle conduit à la détermination des paramètres uniques, servant à décrire le modèle, ce qui n'était pas le cas dans l'approche antérieure. Du fait de la propriété des polynômes orthogonaux, il est donc possible, de partir d'un modèle complexe (modèle saturé par exemple) et rechercher des modèles assez parcimonieux (modèles non saturés) par la suppression de certains termes du modèle de départ sans toutefois réestimer les termes qui en restent. Cette propriété peut permettre d'allouer l'information totale du tableau entre les différentes marges indépendantes de ce tableau. D'ailleurs le principe d'orthogonalisation qu'engendre le lissage par les polynômes orthogonaux rend indépendantes toutes les marges contraignantes du tableau. Dans l'analyse qui suit, nous présenterons d'abord le modèle dans sa forme générale et ensuite, nous construirons deux exemples illustrant l'application du modèle à un tableau de dimension 2 et la généralisation découlera de ces deux exemples.

5.2.2.1. Forme générale du modèle

Soit un tableau de dimension 2, c'est-à-dire un tableau constitué sur la base de deux variables V_1 et V_2 à I_1 et I_2 modalités respectivement. On suppose que ces variables sont ordinales et on assigne aux modalités de ces variables deux ensembles de scores $s_{i_1} (i_1 = 1, \dots, I_1)$ et $s_{i_2} (i_2 = 1, \dots, I_2)$.

Sur cette base, il est possible de construire à partir des scores deux familles de polynômes orthogonaux $P_{i_1} (i_1 = 1, \dots, I_1)$ et $Q_{i_2} (i_2 = 1, \dots, I_2)$ (P_{i_1} est de degré I_1-1 et Q_{i_2} est de degré I_2-1). Chacune des familles est un sous-espace vectoriel de l'espace \mathcal{P} des polynômes orthogonaux de degré n ($I_1 \leq n ; I_2 \leq n$). On cherche donc une base qui engendre les I_1 polynômes orthogonaux P_{i_1} , et une base qui engendre les I_2 polynômes orthogonaux Q_{i_2} . Soient U et V ces bases respectives dont les composantes sont :

$$\begin{aligned} U(i_1) &= P_{i_1}(s_{i_1}); & i_1 &= 1, \dots, I_1 ; I_1=0, 1, \dots, I_1-1 \\ V(i_2) &= Q_{i_2}(s_{i_2}); & i_2 &= 1, \dots, I_2 ; I_2=0, 1, \dots, I_2-1 \end{aligned} \quad (\text{E 5.45})$$

U forme une base orthonormale de \mathcal{R} de même que V forme une base orthonormale de \mathcal{R}

Il est possible de construire une base doublement orthogonale X et permettant de reparamétriser le tableau de dimension 2 à $I_1 \times I_2$ cellules. On a alors :

$$X = U \otimes V \quad (\text{E 5.46})$$

où \otimes est le produit de Kroenecker

Comme il s'agit du modèle Log-linéaire qu'on exprime par le biais des polynômes orthogonaux, en respectant la forme multiplicative du modèle, on peut écrire :

$$\text{Log } t = X \alpha \quad (\text{E 5.47.})$$

où t est bien évidemment le vecteur $(I_1 \times I_2, 1)$ lexicographique des observations du tableau et α le vecteur des paramètres. X est une base orthonormale de. Le modèle (E 5.47) s'écrit encore sous la forme suivante :

$$\text{Log } t(i_1, i_2) = \sum_{i_1}^{I_1-1} \sum_{i_2}^{I_2-1} \alpha_{i_1 i_2} * u(i_1) \otimes v(i_2) \quad (\text{E 5.48})$$

On remarque que ce dernier modèle est presque le même que celui rencontré en (E 3.32) (cf. chap.3), à la seule différence qu'ici, nous avons procédé à une transformation logarithmique des observations du tableau avant de les décomposer par les doubles polynômes orthogonaux. De même, alors qu'en (E 3.32) le modèle était purement explicatif, on peut dire qu'il est ici "descripto-explicatif", ce qui nous amène à penser que les méthodes des régressions par doubles polynômes orthogonaux sont par extension des modèles Log-linéaire (on utilise donc la transformation logarithmique sur les variables qualitatives). Au-delà de cette forme générale, on montrera deux exemples simples de ce modèle appliqués à un tableau de dimension 2.

3.2.2.2. Quelques exemples de modèles illustratifs

Nous présentons ici deux exemples de modèles Log-linéaires fondés sur les polynômes orthogonaux et qui découlent de (E 5.48). Ce sera le cas d'un modèle appliqué au tableau 2x2 (toutes les deux variables sont dichotomiques) et celui d'un modèle appliqué au tableau 3x3 (les variables sont toutes deux trichotomiques), les variables multiclassant ces tableaux étant supposées évidemment ordinales.

5.2.2.2.1 Cas d'un tableau 2x2

On assigne aux modalités de V_1 les scores -1 et 1. Ce même ensemble de scores est assigné aux modalités de V_2 . La recherche d'une base orthonormale U des polynômes orthogonaux P_i ($i_1 = 0,1$) se fera suivant :

$$U = (u(0), u(1)) = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

De même une base orthonormale V des polynômes orthogonaux Q_2 ($I_2 = 0,1$) est telle que:

$$V = ((v(0), v(1)) = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

On trouve ainsi une base orthonormale X associée au tableau 2x2 et définie par :

$$X = U \otimes V = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

de telle sorte que :

$$\text{Logit} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha(0,0) \\ \alpha(0,1) \\ \alpha(1,0) \\ \alpha(1,1) \end{bmatrix} \quad (\text{E 5.49})$$

$\alpha(0, 0)$ est l'effet moyen

$\alpha(1, 0)$ est l'effet linéaire sur la première variable

$\alpha(0, 1)$ est l'effet linéaire sur la deuxième variable

$\alpha(1, 1)$ est l'effet d'interaction linéaire-linéaire entre les deux variables.

Les trois derniers effets sont analogues aux termes d'association des modèles E 5.12), (E 5.11) et (E 5.16) respectivement, du moins dans la façon dont ils s'interprètent.

La matrice X étant inversible, la résolution de (E 5.49) conduit à la détermination des effets uniques et il est donc clair qu'aucune contrainte n'est à imposer aux paramètres. Le modèle des doubles polynômes orthogonaux permet également de réparamétriser un tableau 3x3.

3.2.2.2.2 Cas d'un tableau 3x3

Lorsque le tableau d'étude est un tableau 3x3, on assigne cette fois-ci aux modalités de V_1 les scores -1, 0, 1 et aux modalités de V_2 , on assigne toujours le même ensemble de scores, ce dans le même ordre V_1 et V_2 étant toujours ordinales). Une base orthonormale des polynômes P_i ($i_1 = 0,1,2$) est :

$$U = (u(0), u(1), u(2)) = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 0 & -2 \\ 1 & 1 & 1 \end{bmatrix}$$

V est également une base orthonormale de Q_{i_2} et se définit par :

$$V = (v(0), v(1), v(2)) = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 0 & -2 \\ 1 & 1 & 1 \end{bmatrix}$$

On trouve encore X en effectuant le produit de Kroenecker et le modèle se traduit par :

$$\log t = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & -2 & 0 & -2 & 0 & -2 \\ 1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 0 & -2 & 1 & 1 & 0 & 0 & -2 & -2 \\ 1 & 0 & -2 & 0 & -2 & 0 & 0 & 0 & 4 \\ 1 & 0 & -2 & -1 & 1 & 0 & 0 & 2 & -2 \\ 1 & -1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 0 & -2 & 0 & -2 & 0 & -2 \\ 1 & -1 & 1 & -1 & 1 & 1 & 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \alpha(0,0) \\ \alpha(0,1) \\ \alpha(0,2) \\ \alpha(1,0) \\ \alpha(2,0) \\ \alpha(1,1) \\ \alpha(1,2) \\ \alpha(2,1) \\ \alpha(2,2) \end{bmatrix}$$

Les effets $\alpha(0,0)$, $\alpha(1,0)$, $\alpha(0,1)$ et $\alpha(1,1)$ ont la même interprétation que précédemment.

$\alpha(0,0)$ représente l'effet quadratique sur la deuxième variable.

$\alpha(2,0)$ représente l'effet quadratique sur la première variable.

$\alpha(2, 1)$ est l'effet Quadratique sur la première variable et linéaire sur la deuxième variable

$\alpha(1,2)$ est l'effet linéaire sur la première variable et quadratique sur la deuxième variable.

$\alpha(2, 2)$ est l'effet quadratique sur la première variable et quadratique sur la deuxième variable.

Ici encore, la matrice X est inversible et les effets de paramètres se calculent toujours de façon unique. Lorsque $\alpha(1, 2)$, $\alpha(2, 1)$, $\alpha(1,2)$ et $\alpha(2, 2)$ sont tous nuls, on retrouve le modèle d'indépendance.

Les exemples précédents se généralisent aux tableaux de dimension 2 avec des variables ayant plus de trois modalités et chaque fois il sera toujours possible de trouver une base orthogonale X, du tableau. D'autre part, ces résultats s'étendent également à un tableau de dimension p ($p > 2$) et les effets d'interaction entre les variables prises deux à deux permettent de caractériser l'interaction partielle moyenne entre elles, connaissant toutes les autres. Le seul problème qui risque de se poser est celui de la détermination de X. Dans le cas

des scores équidistants, on détermine les vecteurs U et V à partir de la table des polynômes orthogonaux et par conséquent, X se construit facilement. A l'inverse, si les scores ne sont pas équidistants, la construction de X pose un problème susceptible d'être néanmoins résolu à partir de certains logiciels (c'est le cas de S A S). Toutefois la résolution de ce problème n'exclue pas l'existence d'un autre qui est celui de l'inversion de car cette matrice peut être de taille très grande. A titre d'exemple, un tableau (p,p) aura pour X une matrice (p^2,p^2) (un tableau $(10,10)$ donne une matrice $X(100,100)$) ce qui est énorme. Il faudra alors des logiciels puissants pour l'estimation de paramètres.

Les différents modèles que nous venons de voir dans les paragraphes qui précèdent utilisent dans leur décomposition des scores (modèles Log-linéaires, modèles Logit, doubles polynômes orthogonaux) arbitrairement assignés aux modalités des variables ordinales d'un tableau sur lequel on travaille. Le fondement de cet arbitrage ne semble pas justifié, ce d'autant plus que l'opération d'assignation des scores se fait totalement en dehors des données du tableau. D'autre part, lors de cette opération, l'hypothèse qui prédomine en général est celle d'une équidistance entre les modalités (on ne le dit pas toujours explicitement). Est-il donc vraisemblable de penser à l'issue d'une enquête d'opinion que les modalités "pas du tout favorable", "peu favorable", "assez favorable", "favorable", "très favorable", soient situées dans cette échelle et que la distance séparant deux modalités adjacentes soit identique quelles que soient les modalités ? Même si cette échelle est par essence justifiée, cette assignation a pour conséquence " d'exogénéiser " les scores, c'est-à-dire que ceux-ci ne sont pas obtenus à partir des données et par conséquent, ils ne pourront dans certains cas refléter que partiellement la structure des données du tableau. Il serait idéal de s'ouvrir d'autres voies permettant "d'endogénéiser" les scores . Nous retiendrons celle que nous offre le modèle Log-multiplicatif (BREEN, 1985) nu Log-non linéaire (GOODMAN, 1979).

5.2-3 Forme Log-non linéaire des modèles d'association

La présentation du modèle Log-non linéaire sera succincte, l'une des raisons étant que les différents aspects de ce modèle se retrouvent dans les cas particuliers que sont les modèles (E 5.10) à (E 5.32) où certains scores sont connus a priori, alors qu'ils restent indéfinis dans le cas présent.

Dans ce sous-paragraphe, quelques généralités et propriétés relatives au modèle Log-non linéaire seront décrites avant d'aborder ensuite les procédures d'estimation conduisant à l'ajustement des paramètres de ce modèle. Enfin, une tentative d'utilisation du modèle Log-non linéaire dans l'optique de la technique d'analyse factorielle des correspondances sera évoquée.

5.2.3.1 Caractéristiques du modèle Log-non linéaire

5.2-3-1.1 Quelques généralités sur le modèle Log-non linéaire

Si le tableau d'étude est un tableau de dimension 2, avec ses deux variables ordinales, la forme que prend le modèle d'association est telle que :

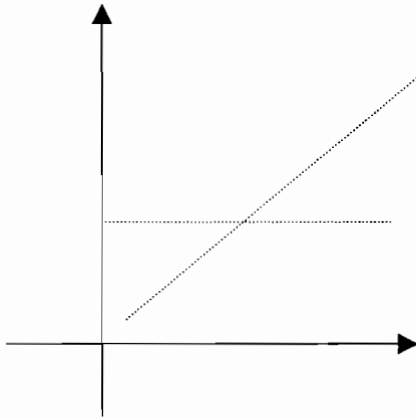
$$t^*(i_1, i_2) = \lambda \lambda_1(i_1) * \lambda_2(i_2) e^{\gamma s_{i_1} s_{i_2}} \quad (\text{E 5.51})$$

S_{i_1} ($i_1=1, \dots, I_1$) et S_{i_2} ($i_2=1, \dots, I_2$) étant des ensembles de scores liés aux modalités des variables V_1 et V_2 respectivement. Les modèles (E 5.10) à (E 5.20) sont les différentes variantes de (E 5.51) par affectation soit de deux sous-ensembles de scores arbitraires aux modalités de V_1 et V_2 , soit l'un des sous-ensembles étant arbitraire, l'autre est estimé à partir du modèle. Bien évidemment dans tous les deux cas, on retrouve la forme Log-linéaire des modèles d'association dans un tableau bidimensionnel. D'ailleurs, le modèle E 5.51) est encore équivalent au modèle (E.5.52).

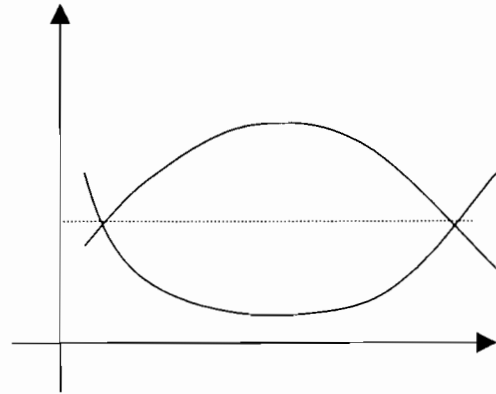
$$\begin{aligned} \text{Log} t^*(i_1, i_2) &= \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \gamma s_{i_1} s_{i_2} \\ \alpha &= \log \lambda \quad \alpha_1(i_1) = \log \lambda_1(i_1) \quad \alpha_2(i_2) = \log \lambda_2(i_2) \quad (\text{E 5.52}) \end{aligned}$$

L'assignation arbitraire des scores aux modalités d'une des variables donne au terme $\gamma s_{i_1} s_{i_2}$ une forme linéaire qui n'existe plus lorsque les deux ensembles de scores sont simultanément inconnus. Dans ce dernier cas, ce terme est de la nature d'une courbure, ce qui confère au modèle la dénomination Log-non linéaire (la décomposition du modèle n'est plus seulement linéaire des paramètres).

Le graphique 4 est l'illustration de la forme de l'association dans le modèle Log-non linéaire. L'association est représentée en ordonnée alors qu'en abscisse sont classées les modalités d'une certaine variable. La concavité de la courbe traduit la positivité de l'association alors que sa convexité est le reflet de la négativité de l'association.



Graphique 3



Graphique 4

En terme de profil, si l'on est en présence d'un tableau de dimension 2, où les modalités -lignes sont par exemple ordonnées, la recherche des scores-colonnes consiste à trouver des grandeurs qui, combinées aux cellules du tableau, permettront d'arriver au profil moyen représenté en pointillé sur le graphique.

Le graphique 3 par contre met en relief la mesure de l'association par le modèle Log-linéaire et on constate d'ailleurs qu'elle prend la forme linéaire. Même en présence d'un tableau de dimension 3, lorsque deux des variables du tableau sont ordinales avec des scores inconnus, on retrouve le modèle Log-non linéaire. Une première façon de rechercher (estimer) ces scores serait de concevoir un modèle d'association "linéaire-linéaire homogène", c'est-à-dire un modèle dans lequel les scores décrivant l'association restent identiques à travers les modalités de la variable restante qui peut être nominale ou ordinale (avec des scores fixés d'avance. Ce modèle sera caractérisé par la recherche des scores $s_{i_1 i_2}$ et $s_{i_1 i_3}$ vérifiant :

$$\begin{aligned}
 s_{i_1} &= s_{i_1} + \\
 s_{i_2 i_1} &= s_{i_2} + & i_1 = 1, \dots, I_1 \\
 & & i_2 = 1, \dots, I_2 \\
 & & i_3 = 1, \dots, I_3
 \end{aligned}
 \quad (E 5.53)$$

Lorsque toutes les trois variables du tableau sont ordinales avec des scores inconnus, alors le modèle Log-non linéaire peut être explicité grâce aux différentes associations partielles, ce par l'intermédiaire des odd-ratios conditionnels.

Ainsi, on a finalement :

$$\text{Log}\theta_{12.3}(i_1, i_2, i_3) = \gamma_{12}(s_{i_1+1} - s_{i_1})(s_{i_2+1} - s_{i_2}) \quad (\text{a})$$

$$\text{Log}\theta_{13.2}(i_1, i_2, i_3) = \gamma_{13}(s_{i_1+1} - s_{i_1})(s_{i_3+1} - s_{i_3}) \quad (\text{b})$$

$$\text{Log}\theta_{1.23}(i_1, i_2, i_3) = \gamma_{23}(s_{i_2+1} - s_{i_2})(s_{i_3+1} - s_{i_3}) \quad (\text{c}) \quad (\text{E 5.54})$$

De ce modèle $S_{i_1}(i_1=1, \dots, I_1)$ est identique dans (a) et (b) de même que le sont $S_{i_2}(i_2=1, \dots, I_2)$ dans (a) et (c) et $S_{i_3}(i_3=1, \dots, I_3)$ dans (b) et (c). C'est donc un modèle dans lequel les différentes associations partielles linéaire-linéaire sont homogènes alors qu'un modèle plus général supposera ces différents scores non identiques d'une association partielle à l'autre (cas du modèle "d'association linéaire-linéaire hétérogène"). Cette approche du modèle Log-non linéaire dans un tableau de dimension 3 peut facilement se généraliser dans un tableau de dimension plus grande de la même manière qu'avec le modèle Log-linéaire.

Le modèle Log-non linéaire est doté de quelques propriétés pouvant être d'un très grand intérêt.

5.2.3.1.2 Propriétés du modèle Log-non linéaire

Le modèle Log-non linéaire a deux principales propriétés, notamment l'invariance de la valeur X^2 du Khi-deux et aussi celle du ratio de la distance relative.

Invariance de la Statistique X^2 de Pearson

Lorsque l'on permute les modalités de V_1 et celles de V_2 , alors la valeur du Khi-deux X^2 associée au modèle log-non linéaire ne change pas (cf. Goodman, 1979, p. 547). Cette propriété n'est pas valable pour les modèles d'association sous la forme log-linéaire, exception faite du modèle d'association nulle modèle d'indépendance). Même en permutant uniquement les modalités de V_1 , ou uniquement celles de V_2 , la valeur du Khi-deux reste également invariante avec le modèle (E5.52).

Le modèle Log-non linéaire peut donc, par l'intermédiaire de certaines permutations conduire à l'ordonnement des modalités des variables lorsque cet ordre n'est pas à priori connu. Dans un tableau $I_1 \times I_2$ les modalités de V_1 et V_2 peuvent être ainsi ordonnées de manière à ce que les scores qui leur correspondent croissent de façon monotone. Un tel cas de

figure mène à la propriété "isotropique" des observations du tableau c'est-à-dire un tableau caractérisé par $\theta_{12}(i_1, i_2)$ (on rappelle qu'il s'agit du odd-ratio) défini par :

$$\theta_{12}(i_1, i_2) \geq 1 \quad i_1=1, \dots, I_1-1 \quad i_2=1, \dots, I_2-1 \quad (E555)$$

La propriété isotropique permettrait ainsi de confronter l'ordre que l'on se donne a priori et celui qui est latent aux modalités. Donc cette propriété a pour mérite d'effectuer un ordonnancement quand celui-ci n'est pas connu au départ.

En revenant au odd-ratio du deuxième ordre auquel on soumet (E 5.52), pour des modalités i_1^1 et i_2^2 de V_1 et i_2^1 et i_1^2 de V_2 , on retrouve encore :

$$\text{Log} \theta_{12}(i_1, i_2) = \gamma_{12}(s_{i_1^1} - s_{i_1^2})(s_{i_2^2} - s_{i_2^1})$$

Toute permutation des modalités de V_1 et V_2 laisse invariante la valeur absolue de γ . Une autre propriété du modèle Log-non linéaire est celle de l'invariance du ratio de la distance relative.

Invariance du ratio de distance relative

La seconde propriété du modèle log-non linéaire est celle concernant les ratios de distance (distance relative). Selon Clogg (1982), "seuls les ratios" de distance entre les scores ont une importance pour le modèle" (modèle Log-non linéaire, bien entendu). En centrant les scores (ne pas confondre avec le centrage des paramètres du chap. 4. il s'agit ici d'un écart par rapport à une constante qui peut par exemple être la moyenne (cf. les modèles (E 5.10) à (E 5.30)) ou en les normant (division des scores par une constante), ces opérations n'altèrent pas la distance relative et par conséquent laissent invariant le modèle Log-non linéaire.

De plus, si l'on permute certaines modalités des variables, la distance relative reste invariante et cette dernière propriété est l'une des grandes caractéristiques du modèle log-non linéaire. On va l'illustrer par un exemple tiré d'un tableau croisant des individus (femmes) suivant leur comportement vis-à-vis de l'avortement et leurs "relations sexuelles prémaritales". La variable "relations sexuelles prémaritales" a quatre modalités et elle est ordinale. On cherche ici à mesurer l'influence des relations sexuelles prémaritales sur le comportement vis-à-vis de l'avortement grâce au modèle log-non linéaire. L'estimation des scores de cette variable donne les valeurs successives :

$$S_1 = -0.280 \quad S_2 = -0.048 \quad S_3 = 0.102 \quad S_4 = 0.225$$

On constate d'abord qu'ils sont monotones croissants. D'autre part, soient d_1 et d_2 les distances absolues telles que:

$$d_1 = S_3 - S_2 \quad d_2 = S_4 - S_3$$

$$d = d_2/d_1 = 0.123/0.150 = 0.82$$

$$d = \text{ratio de la distance relative} \quad (\text{E 5.56a})$$

La permutation des modalités 3 et 4 de la variable "relations sexuelles prémaritales" conduit au nouvel ensemble de scores :

$$S_1^* = -0.448 \quad S_2^* = -0.076 \quad S_3^* = 0.164 \quad S_4^* = 0.361$$

$$\text{On retrouve } d^* = \frac{d_2^*}{d_1^*} = \frac{0.197}{0.240} = 0.82$$

$$\text{Où } d_1^* = S_3^* - S_2^* \quad d_2^* = S_4^* - S_3^* \quad (\text{E3 16b})$$

Il ressort que la permutation des modalités 3 et 4 de la variable "relations sexuelles prémaritales" a induit, au second ensemble de scores un facteur d'échelle de 0,160 par rapport à l'ensemble initial.

Nous quittons le cadre de ces propriétés pour aborder la sous-section relative à la procédure d'estimation dans les modèles d'association en général. Le problème se pose moins pour l'estimation des paramètres du modèle d'association Log-linéaire que celles des paramètres du modèle Log-non linéaire.

5.2.3.2 Estimation des modèles d'association

Les techniques d'estimation susceptibles d'être appliquées au modèle log-linéaire de base (cf. 4.2.) peuvent être transposées aux modèles d'association. Cependant, même si le principe de ces techniques est le même, il y a néanmoins quelques différences, car dans le cas des modèles d'association, l'estimation requiert des contraintes supplémentaires - ce sont des contraintes de marges intégrant l'information relative à l'ordre - qu'il faudra prendre en compte dans la procédure d'estimation. Ici, nous distinguerons l'estimation des modèles d'association log-linéaire, de celle du modèle log-non linéaire qui, comme précisé déjà, présente plus de difficulté.

5.2.3.2.1 Estimation de la forme Log-linéaire des modèles d'association

Les paramètres du modèle d'association Log-linéaire peuvent être estimés par les algorithmes d'ajustement itératif. Deux de ces algorithmes seront présentés ici, et ils ont la particularité d'emprunter l'un et l'autre des chemins totalement opposés: l'algorithme d'ajustement itératif de Simon et Fienberg qui permet d'estimer le tableau des observations avant d'effectuer la décomposition du modèle - on lisse le tableau de façon à ce qu'il satisfasse non seulement les contraintes de marges (contraintes de BIRCH), mais également les contraintes supplémentaires (cf. E5.57) et l'algorithme de GOODMAN (1979) qui, quant à lui, permet d'estimer directement les différents paramètres du modèle. Cet algorithme a l'avantage de converger plus rapidement que le premier

Si l'on veut estimer par exemple le modèle (E 5.11), alors les contraintes de marges à retenir dans le modèle sont les suivantes :

$$\begin{aligned} \sum t^*(i_2, i_2) &= \sum t(i_1, i_2) & i_1 &= 1, \dots, I_1 \\ \sum t^*(i_2, i_2) &= \sum t(i_1, i_2) & i_2 &= 1, \dots, I_2 \end{aligned}$$

A ces contraintes liées bien évidemment aux effets de premier ordre $\alpha_1(i_1)$ et $\alpha_2(i_2)$ on ajoute la contrainte supplémentaire :

$$\sum s_{i_1} t^*(i_1, i_2) = \sum s_{i_1} t(i_1, i_2) \quad (\text{E 5.57})$$

où S_{i_1} ($i_1 = 1, \dots, I_1$) est bien l'ensemble des scores associés aux modalités de la variable ordinaire V_1 .

L'algorithme de Simon et Fienberg appliqué à (E 5.11) requiert les étapes suivantes dans un cycle de l'itération :

$$\begin{aligned} t^*_{(s+1)}(i_1, i_2) &= t^*_{(s)}(i_1, i_2) * \frac{t(i_1, +)}{t^*_{(s)}(i_1, +)} \\ t^*_{(s+2)}(i_1, i_2) &= t^*_{(s+1)}(i_1, i_2) * \frac{t(+, i_2)}{t^*_{(s+1)}(+, i_2)} \\ t^*_{(s+2)}(i_1, i_2) &= \left[\frac{\sum_1 s_1^* t(1, i_2)}{\sum_1 s_1^* t^*_{(s+2)}(1, i_2)} \right]^{s_1^*} * \left[\frac{\sum_1 (1-s_1^*) t(1, i_2)}{\sum_1 (1-s_1^*) t^*_{(s+2)}(1, i_2)} \right]^{1-s_1^*} \end{aligned}$$

$$i=1, \dots, I_1$$

où s_i^* est une transformation de s_i de telle sorte que l'on ait $0 \leq s_i^* \leq 1$ et où $(s=0,1,\dots)$ représente le cycle de l'itération. La relative lenteur de la convergence de cet algorithme sera davantage perçue si l'estimation portait non sur le modèle (E 5.11) mais plutôt sur le modèle d'association uniforme (situation où les deux ensembles de scores sont connus). Dans ce dernier cas, les contraintes (E 5.57) sont remplacées par:

$$\sum s_i s_{i_2} t^*(i_1, i_2) = \sum s_i s_{i_2} t(i_1, i_2) \quad (\text{E 5.59})$$

et à la troisième phase d'un cycle de (E 5.58) est substituée l'étape

$$t^*_{(s+3)}(i_1, i_2) = \left[\frac{\sum_i \sum_j s_i^* s_j^* t(i, j)}{\sum_i \sum_{i_1} s_i^* s_{i_1}^* t^*_{(s+2)}(i, i_2)} \right]^{s_{i_1}^* s_{i_2}^*} * \left[\frac{\sum_i \sum_j (1 - s_i^* s_j^*) t(i, j)}{\sum_i \sum_j (1 - s_i^* s_j^*) t^*_{(s+2)}(i, j)} \right]^{1 - s_{i_1}^* s_{i_2}^*}$$

$$i = 1, \dots, I_1$$

$$j = 1, \dots, I_2$$

$$0 \leq s_i^* \leq 1 \quad 0 \leq s_{i_2}^* \leq 1 \quad (\text{E 5.60})$$

Il faut noter que les modalités de V_1 et V_2 ont des rôles symétriques. Par conséquent si V_2 était ordinale au lieu de V_1 dans la troisième étape de (E 5.58), les modalités i_2 varieraient de telle sorte qu'apparaissent uniquement les valeurs transformées $s_{i_2}^*$ ($0 \leq s_{i_2}^* \leq 1$) comme puissances.

A l'inverse de cette procédure, celle de Goodman qui est également une procédure itérative, appliquée à (E 5.11) donne en partant des solutions de base $\alpha_{1(0)}(i_1)$, $\alpha_{2(0)}(i_2)$ et $\alpha'_{2(0)}(i_2)$ les étapes suivantes au cours de chaque cycle

$$\alpha_{1(s+1)}(i_1) = \alpha_{1(s)}(i_1) * \frac{t(i_1, +)}{t^*_{(s)}(i_1, +)}$$

$$\alpha_{2(s+2)}(i_2) = \alpha_{2(s)}(i_2) * \frac{t(+, i_2)}{t^*_{(s)}(+, i_2)}$$

$$\alpha'_{2(s+3)}(i_2) = \alpha'_{2(s)}(i_2) \left\{ \left[\frac{\sum n_{i_1} [t(i_1, i_2) - t^*_{(s+2)}(i_1, i_2)]}{\sum n_{i_1}^2 t^*_{(s+2)}(i_1, i_2)} \right] \right\} \quad (\text{E 5.61})$$

$n_{i_1} = i_1 - \frac{I_1 + 1}{2}$ c'est-à-dire les scores centrés.

Pour tous les autres modèles d'association Log-linéaires, seule la troisième étape de (E 5.61) change. C'est ainsi qu'avec le modèle (E 5.12) par exemple, cette étape est remplacée par :

$$\alpha'_{1(s+3)}(i_1) = \alpha'_{1(s)}(i_1) \left\{ \left[\frac{\sum n_{i_1} [t(i_1, i_2) - t^*_{(s+2)}(i_1, i_2)]}{\sum n_{i_1}^2 t^*_{(s+2)}(i_1, i_2)} \right] \right\}$$

$$n_{i_1} = i_2 - \frac{I_2 + 1}{2}$$

Ces deux algorithmes restent valables pour l'estimation des paramètres des modèles d'association appliqués à un tableau de dimension 3, voire de dimension supérieure et que nous avons préféré ne pas présenter ici. Une fois ces algorithmes développés, on peut se demander s'ils s'adaptent également à l'estimation du modèle Log-non linéaire qu'on aborde maintenant.

5.2-3.2.2 Estimation des modèles Log-non linéaires

Le présent modèle est caractérisé, on le rappelle, par la non connaissance des scores. Il n'est donc pas possible de lui appliquer l'algorithme itératif de SIMON et FIENERG, puisqu'une telle procédure consisterait à itérer à la fois sur les scores et les observations du tableau simultanément.

L'algorithme de GOODMAN précédemment décrit a été surtout conçu pour l'estimation du modèle Log-non linéaire. Deux autres contraintes seront imposées et nécessaires à cette procédure. Ce sont les contraintes de normalisation : les scores S_{i1} et S_{i2} devront vérifier :

$$\sum s_{i_1} = \sum s_{i_2} = 0$$

$$\sum s_{i_1}^2 = \sum s_{i_2}^2 = 1 \quad (\text{E 5.62})$$

La première contrainte suppose que la somme des scores (c'est une espérance non pondérée) soit nulle. On peut regretter que la deuxième contrainte ne corresponde pas à la nature multiplicative du modèle, puisque dans ce cas, la normalisation en terme d'une somme de carrés est plus compatible avec les outils de la statistique traditionnelle. Il est tout de même possible d'opérer des transformations susceptibles de surmonter cette difficulté. Nonobstant cette incompatibilité non fort comprometteuse, on arrive à des résultats convergents avec la procédure et le principe de celle-ci sera d'ajouter aux deux premières étapes de (E 5.61) les étapes qui suivent :

$$s_{i_1(s+3)} = s_{i_1(s)} + \frac{\sum n_{i_1(s)} [t(i_1, i_2) - t^*_{(s+2)}(i_1, i_2)]}{\sum n_{i_1(s)}^2 t^*_{(s+2)}(i_1, i_2)}$$

$$s_{i_1(s+1)} = s_{i_1(s)} + \frac{\sum n_{i_1(s)} [t(i_1, i_2) - t^*_{(s+2)}(i_1, i_2)]}{\sum n_{i_1(s)}^2 t^*_{(s+2)}(i_1, i_2)} \quad (E 5.63)$$

$$n_{i_1(s)} = s_{i_1(s)} - \bar{s}_{1(s)} \text{ où } \bar{s}_{1(s)} = \frac{1}{I_1} \sum s_{i_1(s)}$$

$$n_{i_2(s)} = s_{i_2(s)} - \bar{s}_{2(s)} \text{ où } \bar{s}_{2(s)} = \frac{1}{I_2} \sum s_{i_2(s)}$$

où $S_{i1}(S)$ et $S_{i2}(S)$ sont des scores estimés au cycle s de l'itération. On pourra prendre comme valeurs initiales de S_{i1} et S_{i2} les valeurs respectives de n_{i_1} et n_{i_2} définies en (E 5.61) et (E 5.61a)

Le modèle Log-non linéaire peut également être estimé par GLIM (BREEN, 1985), (J. DESSENS, W. JANSEN, 1986) qui utilise la procédure de NEWTON-RAPHSON et consistant à itérer sur les moindres carrés pondérés. Cette procédure reste très complexe.

Nous l'avons toutefois appliquée au tableau suivant:

Tableau 9 :

193	8	2	2
486	89	21	20
235	85	52	37
192	132	90	84

Source: Glim Newsletter, n°10, p. 16

L'algorithme converge au bout de 3 cycles d'itération pour un seuil de convergence de 0,01 et donne les ensembles de scores lignes et colonnes estimés par le modèle.

ITERATION	NORMALISED	STANDARDISED	COLUMN	COLUMN
DEVIANCE	SCALE VALUES	SCALE VALUES	VARIABLE	VALUES
PEARSONS	ROW VARIABLE	ROW VALUES		
1.000 5.093 5.769	1 -0.7419	1 -1.484	1 -0.7979	1 -1.592
2.000 3.786 3.887	2 -0.1400	2 -0.2800	2 -0.5635-01	2 -1.127
3.000 3.781 3.870	3 0.2975	3 0.5950	3 0.4144	3 0.8288
CYCLE DEVIANCE DF	4 0.5841	4 1.169	4 0.4378	4 0.8756
3 3.781 4				

On constate d'autre part que la somme de chaque ensemble de scores est nulle, leurs sommes des carrés sont égales à 1, d'où la vérification des contraintes de normalisation.

Tableau 10 : Valeurs estimées de l'interaction multiplicative

0.590	0.041	0.307	-0.325
0.111	0.008	-0.058	-0.061
-0.237	-0.016	0.124	0.130
-0.466	-0.033	0.242	0.255

Le tableau 10 caractérise une situation couramment rencontrée dans certains tableaux de données : sous l'hypothèse d'indépendance, les observations sont plus que celles espérées dans les régions Nord-Ouest et Sud-Est du tableau alors qu'elles sont moins que celles espérées dans la région Nord-Est et Sud-Ouest du même tableau, ce que décrit parfaitement le modèle Log-multiplicatif. Ces sous-régions du tableau indiquent les pôles où il y a des liaisons (liaisons locales) et celles où elles n'existent pas.

D'autres méthodes d'estimation des modèles d'association sont aussi utilisées, à savoir les moindres carrés pondérés par exemple (GRIZZLE, STARMER et KOCH, 1969) qui servent surtout à ajuster les différents modèles logits "accumulés" développés en (E 5.36). Les moindres carrés pondérés ont l'avantage qui est celui de ne pas procéder par itération. DARROCH et RACKLIF (1972) et HABERMAN (1974) ont eux aussi construit des algorithmes itératifs pour estimer les modèles d'association et de nombreux logiciels tels que GENSTAT, FUNCAT, etc... sont couramment employés pour ajuster ces types de modèles.

Il est important de noter au-delà des différentes techniques d'ajustement parcourues, que le principe de choix entre les modèles d'association ou entre les modèles d'association et les modèles Log-linéaires dans leur formulation de base reste le même que celui développé au chapitre 4 (cf. 4.2), c'est-à-dire qu'il reste fondé sur les statistiques du G_2 , les indices d'ajustement des modèles etc. Avant de clore ce chapitre, nous évoquerons succinctement une piste de recherche, pouvant permettre l'extension des possibilités d'exploitation du modèle log-non linéaire.

3.2.3.3 Ebauche d'une autre approche du modèle Log-non linéaire

Le modèle Log-non linéaire tel que nous venons de le présenter s'utilise en vue d'estimer les scores des modalités des variables ordinales. Il semble néanmoins qu'il existe une relation entre ce modèle et la méthode d'analyse des correspondances. Pour y voir plus clair, nous rappellerons brièvement quelques éléments de l'analyse des correspondances avant de tracer le parallèle avec le modèle log-non linéaire, c'est-à-dire examiner dans quelle mesure ce dernier peut servir à la détermination des axes de facteur telle que procède la méthode d'analyse des correspondances.

5.2.3.3.1 Généralités sur la méthode de l'analyse des correspondances

La méthode d'analyse des correspondances, de même que l'outil d'analyse des données qu'est le modèle log-linéaire, voire Log-non linéaire, s'utilise particulièrement pour l'étude des tableaux de contingence. On se limite ici à un tableau à deux dimensions (c'est l'analyse des correspondances simples par opposition à l'analyse des correspondances multiples appliquée aux tableaux binaires ou multidimensionnels (voir GUTTMAN, 1941)) où selon la terminologie couramment utilisée en A.F.C., l'une représente la dimension "individus" et l'autre la dimension "variables" ou caractéristiques (ici les variables correspondent aux modalités définies dans les modèles log-linéaires et Log-non linéaires).

Le principe de la méthode est que chaque dimension du tableau permet de définir les proximités entre les modalités de l'autre dimension grâce à une mesure de la distance qui est la statistique du Khi-deux. L'idée à la base est qu'un tableau de contingence peut conduire à des représentations sous forme de marges dans deux sous-espaces supplémentaires, les ajustements de ces deux marges étant liés par des relations simples (L. LEBART, JP FENELON, 1979).

Contrairement au modèle Log-linéaire très fortement structuré et s'accommodant parfaitement à l'étude d'un tableau où les associations existantes sont représentées en termes d'effets principaux et d'interaction, la méthode d'A.F.C. ne fait appel à aucune structure mathématique et n'est qu'une représentation graphique - "fonction illustrative" - permettant de décrire de façon plus synthétique les données d'un tableau. Par conséquent, elle n'est pas explicitement liée à un modèle .

Partant ainsi d'un tableau bidimensionnel, la détermination des profils individus et des profils - variables permettra de comparer les points individus et les points - variables entre eux et en même temps les situer par rapport au profil moyen (c'est l'origine des axes de projection). Toute distance entre deux points - individus (points - variable) doit être donc modulée selon la distance à l'origine. Ainsi, les proximités entre les points d'une même dimension signifient similitudes de profils par rapport au profil moyen.

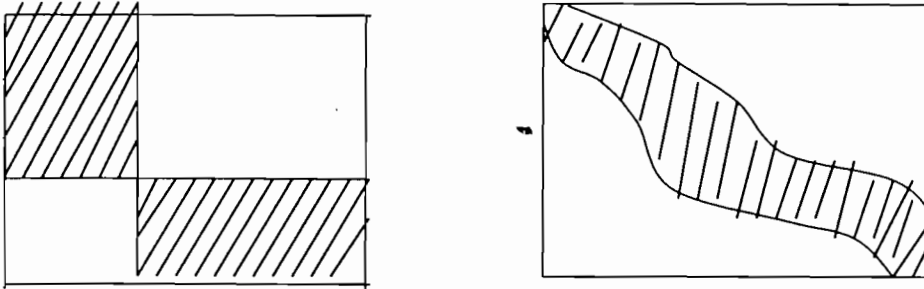
La représentation graphique n'aura d'interprétation fiable que si elle est de meilleure qualité, celle-ci étant mesurée par le pourcentage de la variance totale expliqué par les différents axes de projection. On voit clairement que la méthode repose sur la décomposition de la variance, chaque sous-partie de celle-ci étant associée à un facteur.

Etant donné que l'A.F.C est une méthode graphique, permettant de visualiser possibles entre le,, modalités des variables d'un tableau, elle peut servir à affiner les résultats globaux du modèle Log-linéaire : en effet, le modèle Log-linéaire donne une vue globale des liaisons du tableau, et ne peut détecter les liaisons "locales" à l'intérieur de ce tableau, comme le ferait l'A.F.C. A ce stade, on se demande donc s'il n'est pas possible de concevoir une approche structurelle - non graphique- permettant de prendre en compte dans un tableau la mesure des liaisons "locales". Nous tenterons de l'illustrer grâce au modèle Log-non linéaire tout en signalant au passage qu'on pourra se reporter aux ouvrages de LEBART - FENELON (1979), BENZECRI (1973), HILL (1974) etc.,. pour plus de développement de la méthode d'analyse des correspondances.

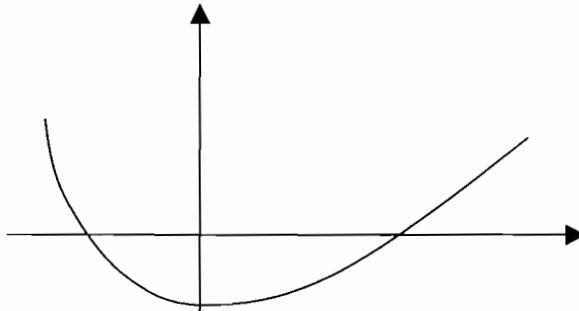
5.2-3-3.2 Approche factorielle du modèle Log-non linéaire

«Lorsque j'ai entendu parler pour la première fois de l'analyse des correspondances (...) il y a eu une discussion très intéressante entre le professeur Escoffier de France qui préconisait l'analyse des correspondances et le professeur Goodman de Chicago qui préconisait quant à lui les modèles logarithmes non linéaires (Log-non linéaires)». Cette citation permet de susciter entièrement l'attention sur le lien étroit pouvant exister entre la méthode d'analyse des correspondances et le modèle Log-non linéaire. En effet, bien que les deux approches aient des fondements différents - l'une est l'illustrative et l'autre formelle - elles ont néanmoins quelques points de similitude : leur faculté à mettre à jour des liaisons locales, c'est-à-dire des liaisons dans une région quelconque d'un tableau.

Soient deux tableaux bidimensionnels définis ci-après :



Ces deux tableaux sont définis chacun par deux sous-ensembles ; le premier a I_1 modalités ($I_1 = I_{11} \cup I_{12}$) et le second en a I_2 ($I_2 = I_{21} \cup I_{22}$). Les régions non hachurées correspondent à celles où les effectifs sont quasi-nuls. L'application de la méthode d'analyse factorielle sur les deux tableaux ne pose pas de problème et dans le cas même du tableau de droite, on verra apparaître l'effet Guttman qui est illustré par le graphique suivant :



Le modèle Log-non linéaire s'adapte parfaitement à l'analyse des deux tableaux et on peut à cet égard se reporter aux résultats que nous procure le tableau 6 (cf. 5.2.3.2). même si traditionnellement, pour détecter des liaisons locales on peut faire recours à la technique d'analyse des résidus : ajustement du modèle d'indépendance et examen des $I_1 * I_2$ résidus. De grands résidus dans ce cas indiquent l'absence d'indépendance et Aitkin pense qu'il est possible d'appliquer à ces résidus la méthode d'analyse des correspondances. L'idée d'Aitkin n'est-elle pas invraisemblable dans la mesure où un tableau des résidus n'est pas un tableau de contingence, ce qui pose un problème quant au fondement de la méthode proprement dite ? Peut-être est-il plus souhaitable dans ce cas d'examiner plutôt les $(I_1 - 1)(I_2 - 1)$ effets d'ordre 2 du modèle saturé parmi lesquels on élimine les effets faibles, les effets restant permettant de

retrouver un modèle de "quasi-indépendance" c'est-à-dire indépendant seulement pour des cellules ayant des interactions faibles ? Il reste au-delà de ces deux voies, que la procédure est longue, contrairement au modèle Log-non linéaire qui décrit directement la structure des tableaux ci-haut.

D'autre part, ce modèle s'apparentant à l'A.F.C., comment pouvons-nous procéder de façon à ce qu'il respecte à la fois la forme multiplicative et la fonction illustrative ? En d'autres termes, comment déterminer les axes des facteurs par le truchement du modèle log-non linéaire ?

L'approche factorielle du modèle Log-non linéaire que nous ébauchons sera fondée sur les statistiques de la théorie de l'information dont J.A. Zighera (1986) a d'ailleurs mis sur pied quelques bases de réflexion. Pour être en conformité avec l'approche factorielle, il est indispensable que soit résolu le problème de la non orthogonalité de l'information totale du tableau, ce problème étant déjà surmonté grâce aux techniques de centrage des paramètres (cf. 4.3.2).

Soient donc t et t^* les vecteurs des valeurs observées et estimées d'un tableau de dimension deux. En guise de rappel, la décomposition de l'information du tableau est définie par :

$$N(t) = N(t/t^*) + N(t^*)$$

où $N(t/t^*)$ est bien évidemment l'information discriminante et $N(t^*)$ l'information du tableau modélisé.

L'opération de centrage des paramètres nous conduit à écrire les modèles d'indépendance et saturé suivant le schéma :

$$\log t^*(i_1, i_2) = \rho_1(i_1) + \rho_2(i_2)$$

$$\log t^*(i_1, i_2) = \rho_1(i_1) + \rho_2(i_2) + \rho_{12}(i_1, 2) \quad (E 5.62)$$

Ces paramètres étant centrés, on calcule les contributions des marges $M(i_1)$ et $M(i_2)$ l'information totale du tableau par :

$$\rho_1(+) = \sum t(i_1, +) \rho_1(i_1)$$

$$\rho_2(+) = \sum t(+, i_2) \rho_2(i_2) \quad (E 5.63)$$

La somme de ces deux grandeurs correspond à $N(t^*)$.

D'autre part, P_{12} qui est l'information résiduelle du tableau est telle que :

$$\rho_{12}(+,+) = \sum \sum t(i_1, i_2) \rho_{12}(i_1, i_2) \quad (\text{E 5.64})$$

cette information résiduelle correspond à $N(t/t^*)$

L'approche factorielle peut être donc conçue comme une opération d'extraction d'un tableau à deux dimensions des effets dus à la première variable et à la deuxième variable - ce sont les effets d'échelle et qui s'identifient à $\rho_1(+)$ et $\rho_2(+)$ dans l'approche informationnelle - laissant le résidu $\rho_{12}(+,+)$ qui peut être interprété comme une succession de profils. Lorsque ceux-ci ne sont pas apparents, il est concevable de les rechercher comme avec la méthode d'analyse des correspondances en définissant d'abord un profil moyen.

Formellement, la recherche du premier axe reviendrait à minimiser l'information discriminante - l'information résiduelle - sous les contraintes :

$$t^*(i_1, +) = t(i_1, +)$$

$$t^*(+, i_2) = (+, i_2)$$

$$\sum s_i t^*(i_1, +) = \sum s_i t(i_1, +) \quad i_1 = 1, \dots, I_1$$

$$\sum s_i t^*(+, i_2) = \sum s_i (+, i_2) \quad i_2 = 1, \dots, I_2$$

En formant le Lagrangien du système et en procédant à sa résolution, alors on retrouve le modèle Log-non linéaire :

$$\text{Log}t^*(i_1, i_2) = \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \gamma s_i s_{i_2}$$

Les valeurs prises par les scores s_{i_1} ou s_{i_2} déterminent le premier axe de la décomposition. D'ailleurs une telle approche reste compatible avec celle de Goodman (1979, 1931) qui a montré que le modèle Log-non linéaire, voire sa généralisation (De Leeuw, 1983, Agriesti, 1983) permettait de déterminer des scores approximativement identiques à ceux fournis sur le premier axe par la méthode d'analyse factorielle des correspondances lorsque les observations issues d'un tableau bidimensionnel suivaient une distribution normale bivariée.

Le premier axe ne représente qu'une partie de l'information résiduelle puisque le centrage a permis d'orthogonaliser cette information. En retirant cette part de l'information

résiduelle, on peut procéder comme précédemment pour chercher à déterminer le second axe qui sera cette fois-ci caractérisé par des scores φ_{i_1} et φ_{i_2} respectivement associés aux modalités des variables V_1 et V_2 . La recherche du second axe nous amènera donc à écrire le modèle suivant :

$$\text{Logit}^*(i_1, i_2) = \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \gamma s_{i_1} s_{i_2} + \theta \varphi_{i_1} \varphi_{i_2}$$

Les décompositions successives peuvent être ainsi opérées de façon à allouer toute l'information résiduelle du tableau aux différents axes de facteurs.

Certes, il ne s'agit là que d'une ébauche de l'approche qui vient d'être effectuée dans ce paragraphe. L'outil informatique que nous disposons actuellement ne nous permet pas encore de tirer des conclusions justifiées sur des résultats éventuels qui découleraient des analyses intégrant le cadre théorique défini ici, et validant pour la circonstance l'approche factorielle du modèle Log-non linéaire. Il semble a priori que ce problème de validité soit levé, étant donné le lien étroit qui existe entre la méthode d'analyse factorielle des correspondances et le modèle Log-non linéaire.

CONCLUSION

Le présent chapitre aura permis d'aborder la possibilité de restituer aux données d'un tableau de contingence ayant des variables ordinales la part de "l'information" relative à l'ordonnement des modalités et omise lorsque l'on traite ce tableau en utilisant la formulation de base du modèle Log-linéaire qui fait totalement abstraction de la nature des variables. Cette restitution s'opère au moyen des différents modèles d'association recensés, desquels se dégagent quelques constatations.

les modèles d'association sont d'une très grande variété et se caractérisent par une structure plus parcimonieuse que celle décrite par les modèles Log-linéaires appliqués aux variables nominales, cette parcimonie étant traduite par l'existence de certains termes d'effets encore plus facilement-interprétables. Quant à la richesse de ces modèles, on remarquera que l'application du modèle Log-linéaire de base à un tableau de dimension 2 par exemple nous conduit soit au modèle d'indépendance, soit au modèle saturé (hypothèse retenue des modèles

hiérarchiques) alors que la prise en compte de l'ordre indique qu'entre les deux modèles, il existe d'autres formes de modèles.

- les modèles d'association sous la forme Log-linéaire et logit ont été construits en faisant implicitement appel à des éléments de la régression simple. Ainsi le paramètre de mesure d'association uniforme peut être saisi comme une pente. En effet, ce paramètre mesure la variation par unité de distance lorsque l'on passe d'un sous-tableau 2x2 à un autre dans un tableau bidimensionnel. De plus les termes d'association sont eux-mêmes linéaires. En somme, on peut dire que les éléments des techniques quantitatives ont été utilisés pour étudier la forme qualitative des modalités ordinales, ce par le mécanisme d'assignation des scores arbitraires à ces modalités.

Au-delà de l'approche factorielle du modèle Log-non linéaire qui a été esquissée, il convient de se garder à l'esprit qu'il s'agit d'une tentative de concilier deux approches. l'une illustrative - la méthode d'analyse des correspondances qui est avant tout une technique graphique et mettant l'accent sur les représentations visuelles - l'autre est un modèle qui appliqué à un tableau permet d'en décrypter les structures. La complémentarité de l'approche illustrative et celle impliquant un modèle (Log-linéaire a été largement mise à contribution dans les travaux de J. DAUDIN, R. TASSONE (1983), P. HEIJDEN et JAN de LEETW (1985) et il sera fort intéressant de tester, lorsque ce sera possible s'il y a un adéquation entre les résultats obtenus par l'approche factorielle du modèle Log-non linéaire et ceux découlant de l'approche complémentaire du modèle Log-linéaire et de la méthode d'analyse factorielle des correspondances.

Outre l'analyse qui vient d'être conduite, une autre voie reste possible : celle qui consiste à saisir la différence entre deux modèles Log-linéaire -le modèle saturé et un modèle quelconque- comme la méthode d'analyse des correspondances (ESCOFFIER 1983, J De LEEUW 1983), pourvu que les paramètres soient centrés, sinon se poserait le problème d'orthogonalisation.

CHAPITRE VI

PRESENTATION DES DONNEES ET ANALYSE DES RESULTATS STATISTIQUES

6.0 INTRODUCTION

Plusieurs approches théoriques ont été abordées dans la première partie du présent travail, qui a permis de développer l'outil d'analyse des données qu'est le modèle Log-linéaire. Ce modèle s'applique à un tableau de contingence, et la construction d'un tel tableau sera effectuée au premier paragraphe (Cf. 6.1), à partir des données des différentes enquêtes Emploi de l'INSEE, qui couvrent une période allant de 1976 à 1981. Ces données sont très riches, et caractérisées par un ensemble de variables servant à les décrire. Quelques unes seulement seront retenues : la profession, la région, l'âge et le sexe, l'objet étant d'analyser l'aspect structurant de l'emploi suivant ces variables. Si l'on fait exception de la variable sexe, le nombre de modalités des trois autres variables est trop important et ne permet pas de mieux exploiter l'information du tableau. Nous simplifierons de ce fait la nomenclature des variables "profession" et "âge" en regroupant certaines de leurs modalités.

Au paragraphe 6.2, nous présenterons quelques résultats statistiques à partir du tableau qui aura été construit en 6.1. Ces résultats vont de la sélection du modèle Log-linéaire aux effets de paramètres aussi bien du modèle Log-linéaire de base que des modèles d'association en passant par les paramètres centrés, c'est-à-dire décorrélés de manière à partitionner l'information totale du tableau entre ses marges indépendantes.

Le paragraphe 6.3 mettra l'accent sur le côté analytique de l'étude. Nous y analyserons les effets de premier ordre associés à chacune des variables du tableau de contingence multidimensionnel sur lequel nous travaillerons.

6.1 PRESENTATION DES DONNEES D'ETUDE

Les données sur lesquelles s'appuie l'étude que nous mènerons ici sont tirées d'une banque de données INSEE, réalisée pour le C.A.S.S.F. et qui reprend en quelque sorte certaines caractéristiques des individus figurant dans les différentes enquêtes Emploi de mars, généralement effectuées par l'INSEE. Ces données sont celles qui couvrent une période allant de 1976 à 1981, et nous évoquerons plus loin pourquoi il n'a pas été possible de prolonger les analyses qui en découlent jusqu'à une période plus récente, bien qu'il existe un fichier de données susceptible d'être exploité au-delà de 1981.

Les Enquêtes Emploi, qui utilisent des échantillons au 1/300^e procurent en général des résultats assez satisfaisants et sont essentiellement centrées sur des individus (on exclue du champ d'enquête les personnes vivant en ménages collectifs), lesquels serviront à construire des tableaux d'effectifs - ce sont des tableaux de contingence - en vue de l'application de l'outil d'analyse développé jusqu'ici (Cf. première partie).

Quelques caractéristiques (variables) ont été tirées de ces enquêtes et permettent de constituer un tableau de contingence multidimensionnel.

- le sexe qui a évidemment deux modalités (homme, femme),
- l'âge reprend toutes les classes d'âge de 0 à 99 ans , cette variable a donc 1 00 modalités,
 - la profession possède 59 modalités que sont les différents postes de la classification 59PJ (cf. Annexe A9).
 - la région est une variable à 22 modalités (les 22 régions françaises)
 - l'année a 6 modalités : 1976, 1977, 1978, 1979, 1980, 1981.

Sur la base de ces 5 variables, il est donc possible de construire un tableau de contingence multidimensionnel $T(2, 100, 59, 22, 6)$ qui a 1 557.600 cellules ($2 \times 100 \times 59 \times 22 \times 6$). Tout étant riche en information, ce tableau présente un inconvénient majeur qui est sa taille très grande, ce qui induit deux types de problèmes :

- celui des possibilités de calcul des paramètres du tableau. Ces possibilités sont limitées dans la mesure où les différents logiciels de traitement que nous utilisons (BMDP, GLIM) ne sont pas capables de restituer des résultats au-delà d'un certain seuil acceptable en ce qui concerne le nombre de cellules du tableau d'étude.

- le deuxième type de problème est le corollaire du premier et est celui de l'accès à l'information du tableau. En effet, les capacités limitées des logiciels impliquent l'impossibilité de détecter toutes les structures latentes d'un tableau à très grand nombre de cellules.

Ces deux problèmes nous amènent à rechercher un tableau multidimensionnel dont la taille des cellules est réduite dans les proportions acceptables par les logiciels de traitement. La recherche d'un tel tableau consistera à agréger certaines modalités des variables du tableau $T(2, 100, 59, 22, 6)$ défini précédemment. Certes, l'opération d'agrégation entraîne en elle-

même une perte d'information du tableau. Mais celle-ci n'est - elle pas mineure au regard du problème d'inaccessibilité à l'information du tableau 7 ?

L'idéal pour nous aurait été d'effectuer des agrégations qui assurent des pertes d'information minimum grâce aux outils de mesure que sont l'information discriminante symétrisée et la perte de déviance ou d'autres algorithmes d'agrégation. Mais, dans la mesure où l'une des finalités de l'analyse que nous envisagerons par la suite reste implicitement de vérifier si certaines études effectuées au C.A.S.S.F par d'autres outils d'analyse étaient compatibles avec la méthodologie qu'on se propose ici, on a préféré conserver les différentes nomenclatures utilisées pour ces précédentes études et dont nous résumons néanmoins les traits principaux de leur réduction, c'est-à-dire la construction des nomenclatures finales des variables profession, âge, région.

6.1.1 Réduction de la nomenclature de la variable profession

Avant de décrire les principaux groupes de métiers, il faut souligner que le critère qui avait été choisi pour l'agrégation était de constituer des groupes homogènes qui atteignent chacun ou approchent au moins 50 000 femmes salariées.

Sur la base de ce critère, il a été possible de ramener la nomenclature initiale en 59 postes à une nomenclature en 23 postes, cette opération s'effectuant en deux temps :

- dans un premier temps, les regroupements des métiers sont sectoriels (agrégation dans un même groupe des métiers d'un même secteur). Neuf grands groupes sont ainsi formés dont un du secteur primaire, quatre du secteur secondaire et quatre du secteur tertiaire (cf. Tableau 10) les groupes de chacun des deux derniers secteurs étant liés à un niveau de qualification (il existe six niveaux de qualification).

- dans un deuxième temps, les groupes de professions plus précis sont formés au sein de chacun des groupes du tableau 10, chaque groupe étant constitué à niveau de qualification identique.

Tableau 10 : Nomenclature simplifiée des professions par secteurs.

METIERS PRIMAIRES	METIERS SECONDAIRES	METIERS TERTIAIRES	NIVEAU
agriculteurs, bûcherons, marins et pêcheurs	ingénieurs, architectes, scientifiques (1)	cadres supérieurs, médecins, professeurs, officiers et juristes (5)	I-II
	Techniciens et dessinateurs (2)	cadres moyens, comptables, métier technique et instituteurs (9)	III-IV
	maîtrise et ouvriers qualifiés (25)	employés qualifiés, secrétariat, métiers qualifiés, armées (9)	V
	ouvriers spécialisés et manœuvres (1)	employés non qualifiés, personnel de service et de surveillance (5)	VI

Les chiffres entre parenthèse représentent le nombre des profession retenues dans le groupe constitué.

En procédant ainsi, 27 groupes de professions ont été constitués (voir nomenclature ci-après).

Nomenclature simplifiée des professions

1. Ouvriers qualifiés du travail des étoffes
2. Métiers de la santé (C.A.P)
3. Personnel diplômé des services paramédicaux et sociaux
4. Métiers des soins personnels
5. Personnel qualifié des techniques comptables
6. Mécanographe, opérateur, perforateur
7. Métiers non qualifiés du commerce et de la distribution
8. Employés non qualifiés
9. Métiers techniques de l'hôtellerie
10. Métiers qualifiés de la vente et de l'hôtellerie
11. Cadres moyens

12. Comptables et techniciens des banques
13. Métiers non qualifiés de l'hôtellerie
14. Professeurs, professions littéraires et de l'information
15. Employés administratifs qualifiés
16. Ouvriers qualifiés des textiles, cuirs et peaux et chimie
17. Ouvriers spécialisés et manœuvres
18. Ouvriers qualifiés de la radio-électricité, du papier Carton et autres spécialités de l'alimentation
19. Métiers techniques de la distribution
20. Cadres administratifs supérieurs
21. Ouvriers qualifiés verre, photographie
22. Techniciens et dessinateurs
23. Autres ouvriers qualifiés
24. Personnel de secrétariat
25. Personnel de service
26. Instituteurs et assimilés
27. Reste.

Le poste 27 (Reste) de la nomenclature simplifiée peut être exclu du tableau car il regroupe à la fois les effectifs de l'agriculture et des métiers à effectifs faibles difficilement agrégeables à l'autre. En quelque sorte le tableau se réduira aux métiers du secondaire et du tertiaire. D'autre part, les postes 24 à 26 sont également exclus du tableau dans la mesure où les études que nous avons évoquées plus haut portaient sur la "participation féminine régionale" dont on sait que pour ces postes, le comportement régional n'est pas sensiblement différent du comportement national. Aussi, une étude qui intègre ces différents postes n'apporte pas d'information substantielle par rapport à celle qui ne les prend pas en compte. On est donc finalement réduit à 23 groupes de professions (les 23 premiers groupes de la nomenclature. Le groupe de profession "Autres ouvriers qualifiés" est le groupe le plus hétérogène, puisqu'il comprend plusieurs types d'ouvriers qualifiés (cf. annexe A.9). Cette réduction de la nomenclature des professions effectuée, il est indispensable d'en faire autant pour les classes d'âges.

6.1.2 Réduction de la nomenclature de la variable âge.

Tout comme la variable profession, la variable âge a été longtemps retenue dans les études du C.A.S.S.F. Nous retiendrons six groupes d'âge à partir des 100 modalités initiales que comporte la variable et qui sont les suivantes :

- groupe 1 : âge < 25 ans
- groupe 2 : 25 ans < âge < 31 ans
- groupe 3 : 32 ans < âge < 39 ans
- groupe 4 : 40 ans < âge < 47 ans
- groupe 5 : 48 ans < âge < 54 ans
- groupe 6 : âge > 54 ans

Ces groupes sont également maintenus afin de confronter certains résultats obtenus au moyen d'autres approches avec ceux qu'on tirerait de l'approche Log-linéaire.

D'autres variables ne subiront pas ou subiront peu de regroupement, soit parce qu'elles ont des modalités qui restent très modestes, soit pour des raisons diverses :

- la variable région passe de 22 modalités à 21 modalités, la région Corse, très petite étant regroupée avec la Provence Alpes Côte d'Azur.

Certes, il aurait été possible de regrouper des régions, mais nous avons souhaité prospecter dans l'analyse ultérieure le comportement spécifique de chaque région suivant certaines caractéristiques. Toute agrégation ne permet pas d'envisager de façon non préjudiciable une telle étude.

Nomenclature des régions

1. Ile-de-France
2. Champagne-Ardenne
3. Picardie
4. Haute-Normandie
5. Centre
6. Basse-Normandie

7. Bourgogne
8. Nord-Pas-de-Calais
9. Lorraine
10. Alsace
11. Franche-Comté
12. Pays de la Loire
13. Bretagne
14. Poitou Charentes
15. Aquitaine
16. Midi-Pyrénées
17. Limousin
18. Rhône-Alpes
19. Auvergne
20. Languedoc-Roussillon
21. Provence - Côte d'Azur - Corse

Ces différentes simplifications étant opérées, il reste à savoir de quelle manière doit être traitée l'année. Faut-il la considérer comme une variable ou non ? Il serait peut être fastidieux d'intégrer dans l'analyse comme variable l'année puisque le tableau multicroisé à constituer avec les nomenclatures précédemment simplifiées aurait une taille encore relativement grande. C'est pourquoi nous mènerons une étude année par année et par conséquent, seront construits six tableaux de contingence $T(21, 23, 6, 2)$ de dimension 4 (21 régions, 23 groupes de professions, 6 groupes d'âge, les 2 modalités de la variable sexe). Chacun des six tableaux aura 5796 cellules. Pour l'instant, on peut s'arrêter à ce niveau de simplification, tout en sachant qu'elle reste insuffisante lorsqu'il faudra exploiter certains éléments de la théorie développés au chapitre 5. Nous y reviendrons quand il sera nécessaire.

Il convient de signaler qu'il aurait été possible de travailler sur les tableaux conçus pour les années 1982 et 1983 puisque les données relatives à ces années nous sont disponibles. Malheureusement, l'exploitation des nomenclatures de tels tableaux poserait des problèmes d'homogénéisation des nomenclatures, dans la mesure où la nomenclature des professions, homogène sur la période 1976-1981 a changé à partir de 1982.

La simplification des nomenclatures et la construction d'un tableau de contingence (on a l'exemple d'un tel tableau à 4 dimensions en annexe A6) nous conduisent à la phase de la modélisation Log-linéaire proprement dite, afin de rechercher la validation des approches théoriques développées dans les chapitres 3, 4, 5. Autrement dit, nous explorerons les résultats statistiques du modèle et l'estimation des paramètres aussi bien du modèle Log-linéaire de base que ceux du modèle Log-linéaire dans son extension voire des modèles d'association .

6.2. EXPLORATION DE QUELQUES RESULTATS STATISTIQUES

Les résultats statistiques à explorer ici sont ceux liés au choix de modèles, à l'estimation des paramètres du modèle retenu (modèle Log-linéaire de base) et à celle des paramètres du modèle Log-linéaire dans son approche extensive (paramètres centrés).

6.2.1 Sélection du meilleur modèle

Pour chacune des années d'étude, nous chercherons le modèle qui permet de mieux ajuster les données du tableau $T(21, 23, 6, 2)$, ce suivant le principe décrit en 4.3 (cf. chapitre 4). L'échantillon ayant permis de constituer le tableau $T(21, 23, 6, 2)$ étant de taille très grande, les différentes statistiques d'ajustement des modèles (voir tableau 11) sont largement significatives. L'accent sera mis ici sur les indices d'ajustement des modèles - indice de GOODMAN - qui indiquent l'amélioration relative d'un modèle lorsque l'on ajoute ou supprime un (des) terme(s) d'effets de ce modèle d'une part et qui corrigent le biais induit par la taille de l'échantillon d'autre part. Nous choisirons un modèle de référence parmi les modèles suivants :

- modèle d'indépendance
- modèle contenant tous les effets de 2^{ème} ordre
- modèle contenant tous les effets de 3^{ème} ordre
- modèle saturé

Au-delà, on essaiera d'améliorer ce modèle de référence par suppression ou inclusion d'autres termes d'effets.

Le tableau 12 permet de mesurer la part d'information (ou encore l'inertie expliquée) que procurent le modèle d'indépendance, le modèle contenant tous les effets de 2^{ème} ordre, le

modèle contenant tous les effets de 3^{ème} ordre et le modèle saturé, Quelle que soit l'année d'étude, on remarque que les différents indices ne varient presque pas.

Tableau 11 : des valeurs du G^2 et des degrés de liberté (ddl) associés à chaque modèle sur la période d'étude

ANNEES MODELES	1976		1977		1978		1979		1980		1981	
	G^2	ddl	G^2	ddl	G^2	ddl	G^2	ddl	G^2	ddl	G^2	ddl
Equire-partition	3613	5795	3614	5795	3589	5795	3505	5795	3536	5795	3440	5795
Effets principaux	926	5747	928	5747	945	5747	967	5747	984	5747	966	5747
Effets d'ordre 2	188	5050	297	5050	199	5050	199	5050	199	5050	195	5050
Effets d'ordre 3	65	2200	67	2200	66	2200	69	2200	72	2200	70	2200
Effets d'ordre 4	0	0	0	0	0	0	0	0	0	0	0	0

Tableau 12 : Variation par rapport à l'équipartition de l'inertie expliquée par les modèles (en %)

ANNEES	1976	1977	1978	1978	1980	1981
M_1	74.4	74.3	73.7	72.4	72.2	71.9
M_2	98.8	94.5	94.4	94.3	94.4	94.3
M_3	98.2	98.1	98.2	98.0	97.9	97.9
M_4	100	100	100	100	100	100

Tableau 13 : Perte d'information due à la suppression d'un terme d'effet d'ordre 2 du modèle M_2 (en %)

ANNEES	1976	1977	1978	1979	1980	1981
M_2 -GR	0.48	0.40	0.35	0.34	0.31	0.36
M_2 -PR	2.60	2.90	2.80	2.70	2.98	3.00
M_2 -SR	0.16	0.15	0.14	0.10	0.11	0.10
M_2 -PS	13.20	12.80	13.40	14.50	14.60	14.90
M_2 -GP	3.25	3.30	3.40	3.60	3.64	3.50
M_2 -GS	0.11	0.10	0.10	0.06	0.03	0.05

Tableau 14 : mesure des liaisons globales (en %)

ANNEES	1976	1977	1978	1979	1980	1981
M_2 - M_1	78.7	78.7	78.9	79.4	79.7	79.7
M_3 - M_1	92.9	92.7	93.0	92.8	92.6	92.7
M_4 - M_1	100	100	100	100	100	100

Sur la période 1976-1981, le tableau 12 montre que le modèle d'indépendance explique en moyenne 73 % de l'inertie totale des données du tableau T(21, 23, 6, 2). Cette part est de 94,5 % pour le modèle contenant tous les effets de 2^{ème} ordre, soit une amélioration de 21,5 % de l'inertie apportée uniquement par tous les effets de 2^{ème} ordre.

Le modèle contenant tous les effets de 3^{ème} ordre explique en moyenne 98 % d'inertie totale des données sur la période d'étude et l'amélioration apportée par tous les effets de 3^{ème} ordre, par rapport au modèle contenant tous les effets de 2^{ème} ordre n'est plus que de 3,5 % en moyenne. De même le modèle saturé qui ajuste totalement les données du tableau est caractérisé par la nullité de sa déviance, ce qui correspond en terme d'inertie à l'explication à

100% de l'inertie totale du tableau. On constatera bien évidemment que la contribution à l'inertie totale de l'interaction d'ordre 4 n'est que de 2 % en moyenne .

Ces différentes inerties guideront le choix de modèle de référence que nous effectuerons. Certes, il est clair que plus interviennent les effets de paramètres plus un modèle est bon. Mais quel serait l'intérêt par exemple de retenir un modèle à structure complexe incluant les interactions difficilement interprétables.

Compte tenu du principe de la parcimonie et du fait que les modèles d'ordre 3 et saturé n'améliorent que peu le modèle incluant tous les effets de deuxième ordre, ce dernier est choisi comme modèle de référence. Il se formalise suivant :

$$\log t(i_1, i_2, i_3, i_4) = \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \alpha_3(i_3) + \alpha_4(i_4) + \alpha_{12}(i_1, i_2) + \alpha_{13}(i_1, i_3) + \alpha_{14}(i_1, i_4) \\ + \alpha_{23}(i_2, i_3) + \alpha_{24}(i_2, i_4) + \alpha_{34}(i_3, i_4) \\ i_1=1, \dots, 21 ; i_2=1, \dots, 23 ; i_3=1, \dots, 6 ; i_4=1, 2$$

avec les contraintes ANOVA.

Ce modèle reste encore perfectible. L'élimination successive des différents termes d'interaction du modèle de référence permet de mesurer l'inertie perdue (la perte d'information) par cette opération, Suit M₂ le modèle de référence, Les différents termes d'interaction $\alpha_{12}(i_1, i_2), \alpha_{13}(i_1, i_3), \alpha_{14}(i_1, i_4), \alpha_{23}(i_2, i_3), \alpha_{24}(i_2, i_4), \alpha_{34}(i_3, i_4)$ correspondent respectivement à RP, RG, RS, PG, PS, GS (Région x Profession, Région x Age, Région x Sexe, Profession x Age, Profession x Sexe , Age x Sexe). L'examen du tableau 13 montre que la suppression du terme RG du modèle M₂ n'altère presque pas la statistique d'ajustement de M₂, puisque pour l'année 1976 par exemple, la perte d'inertie n'est que de 0,48 %. Cette perte d'inertie devient même faible sur la période.

La suppression successive des termes RS et GS de M₂ n'apporte pas également de modification sensible de la statistique d'ajustement de M₂ : les pertes d'inertie par rapport à M₂ ne sont que de 0.16 % et 0,11 % en ce qui concerne RS et GS respectivement, pour l'année 1976. Par contre si l'on élimine de M₂ le terme PS, il y a très forte perte d'inertie de M₂ et elle est de 13,2% pour l'année 1976. Le résultat est encore presque similaire sur la période d'étude.

Les autres termes RP et PG engendrent des pertes d'inertie moindres que PS si on les élimine individuellement de M_2 (2,6 % et 3,25 % respectivement pour l'année 1976), mais plus fortes que les autres termes.

On peut donc dire que les termes PS PR et PG ne sauraient être exclus de M_2 , dans la mesure où leur contribution informationnelle est relativement, plus importante que celle apportée par RS, GS et RG et qui reste négligeable. Ces trois termes peuvent donc être supprimés du modèle M_2 . On vérifiera dans le tableau 15 que la suppression des trois termes de M_2 donne le modèle M_2 dont les indices d'ajustement ne se démarquent presque pas de ceux de M_2 .

La conclusion suivante peut être faite à partir de l'analyse qui vient d'être effectuée : il existe une très forte liaison entre profession et sexe, une forte liaison entre profession et région d'une part et profession et âge d'autre part, des liaisons marginales entre les autres termes d'effets (il s'agit des liaisons conditionnelles), toutes choses égales par ailleurs. Ces différentes liaisons sont illustrées par le, graphique 5.

Ce graphique fait apparaître des liaisons fortes entre la variable profession et chacune des trois autres variables. Ceci est une indication du pouvoir de discrimination relativement plus important de la variable profession par rapport aux autres variables. Nous aurons d'ailleurs l'occasion de le vérifier par la suite (Cf.6.3.1). En somme, le "meilleur" modèle est le suivant:

$$\log t(i_1, i_2, i_3, i_4) = \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \alpha_3(i_3) + \alpha_4(i_4) + \alpha_{12}(i_1, i_2) \\ + \alpha_{23}(i_2, i_3) + \alpha_{24}(i_2, i_4)$$

$$i_1=1,2\dots,21$$

$$i_2=1,2\dots,23$$

$$i_3=1,2\dots,6$$

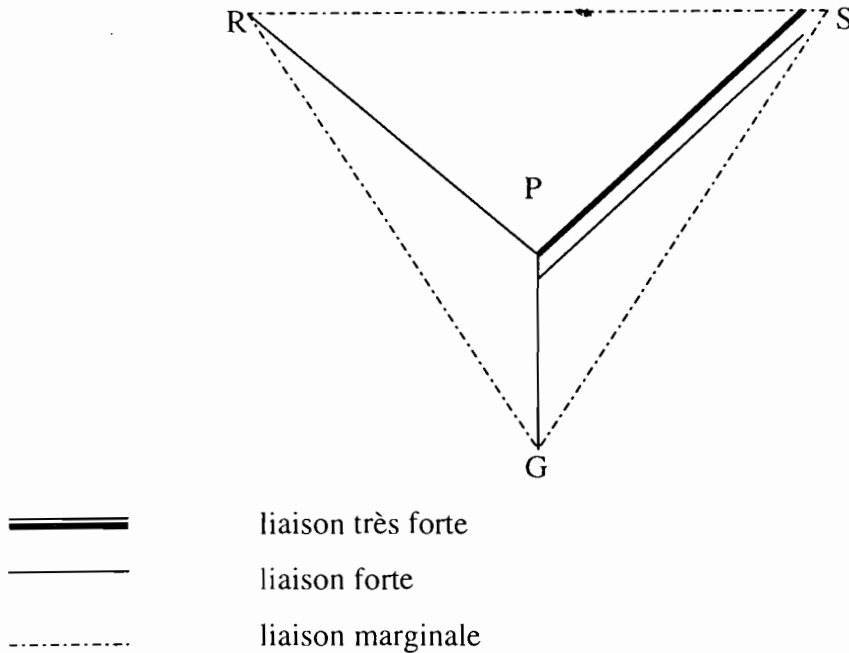
$$i_4=1,2$$

La critique qu'on peut porter à ce modèle est qu'il ne met en relief que des liaisons globales. En effet, le modèle précédent indique simplement que, appliqué au tableau à quatre dimensions défini en (6.1), il existe globalement entre les variables les liaisons conditionnelles suivantes :

- liaison de dépendance conditionnelle entre profession et sexe (la profession est fortement liée au sexe, toutes les régions et toutes les classes d'âge confondues).

- liaison de dépendance conditionnelle entre profession et âge.
- liaison de dépendance conditionnelle entre profession et région.

**Graphique 5 : REPRESENTATIONS DES LIAISONS GLOBALES
ENTRE LES VARIABLES**



Ce modèle est incapable de détecter les liaisons locales entre les variables. Ces liaisons pourraient être induites par certaines modalités des variables qui, globalement ne sont pas saisies peut-être à cause de la sous-représentativité de ces modalités. C'est ainsi que globalement., il n'y a pas de liaison entre région et âge alors que cette liaison peut exister pour certaines régions et/ou certaines classes d'âge.

Le modèle Log-non linéaire aurait pu nous éclairer davantage sur de telles liaisons, mais il est regrettable que nous n'ayons pu l'exploiter à une dimension plus grande pour des raisons qui tiennent de la capacité du logiciel de traitement. Pour l'instant, il n'a été possible de l'appliquer qu'à un tableau de dimension 2, ce qui implique la perte de deux dimensions de notre tableau d'étude et par conséquent l'élimination des différents termes d'effets de paramètres mesurant les liaisons conditionnelles dit modèle. Les recherches entreprises actuellement au C.A.S.S.F permettront sûrement de rendre possible toute exploitation du modèle Log-non linéaire à l'échelle d'un tableau de dimension supérieure à 2.

Outre ces remarques, nous appliquerons le modèle retenu au tableau croisant la profession, la région, l'âge et le sexe afin d'estimer les paramètres qui lissent ce tableau, ne serait-ce qu'à titre indicatif puisque les insuffisances de ce modèle (cf. 4.3) nous invitent à axer l'analyse beaucoup plus sur les paramètres centrés du modèle Log-linéaire.

6.2.2 Effets de paramètres du modèle Log-linéaire

Nous présenterons les effets de paramètres du modèle Log-linéaire de base, ceux du modèle Log-linéaire dans son extension et ceux des modèles d'association.

6.2.2.1. Paramètres du modèle Log-linéaire de base

L'application du logiciel BMDP au tableau T(21, 23, 6, 2) a permis d'obtenir les effets de premier ordre et les effets d'interaction suivants :

R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12
1.914	-0.423	-0.117	-0.107	0.056	-0.518	-0.555	0.663	0.179	-0.183	-6.22	6.260

R13	R14	R15	R16	R17	R18	R19	R20	R21
0.063	-0.449	0.127	0.072	-1.207	0.929	-0.760	-0.412	0.587

P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
-0.551	-0.106	-0.166	-0.993	-0.528	-1.094	-0.703	0.771	-1.432	1.163	0.241	-0.120

P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23
-1.242	0.029	0.545	-0.744	2.142	0.301	0.174	0.625	-1.207	0.863	2.025

A1	A2	A3	A4	A5	A6		S1	S2
0.099	0.283	0.040	0.074	-0.153	-0.342		0.042	-0.042

On vérifiera que les contraintes ANOVA sont respectées, la somme des effets de premier ordre est nulle. De même pour les effets de second ordre, on constate quelle que soit la ligne (respectivement la colonne) que la somme des effets de paramètres est nulle. Comme

on l'a déjà vu sur le plan théorique, les paramètres du modèle Log-linéaire de base ont peu d'intérêt et ne servent qu'à prospecter la structure d'un tableau lorsque l'on procède au lissage de celui-ci. Du fait de la non unicité de ces paramètres, on ne peut leur accorder une signification concrète. En effet, ils appartiennent à un hyperplan et toute combinaison linéaire des paramètres de cet hyperplan peut conduire à la valeur estimée de chaque cellule du tableau. Pour illustrer l'aspect non significatif des paramètres, on peut s'arrêter un instant aux effets liés au sexe. Soit donc l'effet de premier ordre associé au sexe. La valeur que prend la modalité homme est 0,042 (la positivité de cet effet indique une sur-représentativité des hommes par rapport à la moyenne).

Ainsi qu'on le verra ultérieurement les hommes représentent un peu plus de 60 % des effectifs dans l'échantillon total du tableau sur lequel nous avons travaillé. La valeur -0,042 tout en indiquant que les femmes sont sous-représentées, ne reflète pas nécessairement leur poids dans l'échantillon. Cette valeur a été obtenue par nécessité de rendre la somme des effets de paramètres nulle, contrainte indispensable pour leur estimation. D'autre part, les paramètres sont corrélés entre eux (cf. marges impliquées et impliquantes). Afin de les décorréler et les rendre indépendants les uns les autres, nous avons utilisé le modèle Log-linéaire dans son extension et qui donne des paramètres centrés, caractérisés par leur unicité et par conséquent leur significativité.

6.2.2.2 Effets de paramètres centrés

Grâce au programme conçu au C.A.S.S.F* , il est désormais possible de donner un sens aux paramètres du modèle Log-linéaire par procédure de centrage, c'est-à-dire leur orthogonalisation de façon à rendre indépendantes les informations associées aux différentes marges significatives du tableau. L'application de ce programme au tableau T (21, 23, 6, 2) pour l'année 1976 nous donne un certain nombre de résultats qui restent globalement valables pour les autres années. Nous présenterons donc ces résultats en renvoyant ceux des années 1977, 1978, 1979, 1980 et 1981 en annexe.

L'examen des effets simples indique que la variable profession est la plus discriminante. Sa contribution absolue à l'information est de 539 millinats. Viennent ensuite les variables région (324 millinats), sexe (44 millinats) et l'âge (19 millinats).

Tableau 15a : Effet "région" centré.

1660	-679	-372	-371	-203	-776	-517	403	-78	-431	883	2	193	714	-123	187
-1476	671	-824	-680	319	324										
3804	367	498	499	590	333	431	1082	669	470	299	725	796	354	633	690
165	1415	317	366	995	1383										
249	24	33	33	39	22	28	71	44	31	20	48	39	23	42	39
11	93	21	24	66	1000										

Tableau 15b : Effet "profession" centré.

-1073	-679	-688	-1534	-1039	-1603	-1225	264	-1967	629	-276	632	1763	486	33	1255
1632	-215	-331	125	-1725	351	1511	539								
200	311	293	126	206	117	171	760	82	1025	443	310	100	359	600	166
2985	471	419	661	104	829	2644	1714								
15	23	22	10	15	9	13	56	6	83	33	23	7	27	45	12
221	35	31	48	8	61	196	1000								

Tableau 15c : Effet "âge" centré.

86	264	22	55	-173	-368	19
1069	1277	1003	1036	825	679	1019
180	217	170	176	140	116	1000

Tableau 15d : Effet "sexe" centré.

260	-349	44
1240	674	1045
646	354	1000

Chaque tableau a trois lignes :

- la première ligne indique les effets de paramètres associés à chacune des modalités de chaque variable. La dernière colonne de cette ligne est la contribution absolue de la marge correspondante à l'information des données du tableau.

- la deuxième ligne représente la valeur normée des effets de paramètres (la dernière colonne de cette ligne est égale à la valeur normée de la marge associée).

- la troisième ligne est le poids de chacune des modalités de la marge considérée (la dernière colonne indique le poids total du tableau qui a été ramené à 1000).

On vérifie également ici que la moyenne harmonique exponentialisée des paramètres est égale à l'unité.

Il est à signaler que les valeurs prises par les grandeurs à l'intérieur du tableau ont été toutes multipliées par 1000. Donc avant d'envisager toute opération (exponentialisation par exemple) , il est important de prendre les valeurs au 1/1000^e.

La lecture des effets de second ordre se fera de manière analogue à celle des effets de premier ordre. Toutefois on aura affaire à deux marges d'ordre un totalement indépendantes qu'implique la marge d'ordre deux, ce qui fait qu'à la dernière colonne de la ligne i_1 , on a la contribution absolue de la modalité i_1 de la première variable ; à la dernière ligne se trouvent les contributions absolues des modalités i_2 de la deuxième variable. A l'intérieur du tableau, on a les valeurs des paramètres (valeurs centrées) , leurs valeurs normées et leurs poids aux premières, deuxièmes et troisièmes lignes respectivement.

La valeur à la marge indique la contribution du tableau à l'information totale des données.

Le tableau profession x sexe a la plus grande contribution informationnelle (169 millinats). Cette contribution est de 42 millinats pour le tableau profession x Age, 35 millinats pour le tableau profession x Région. Les tableaux Région x Age, Région x Sexe et Age x Sexe ont respectivement comme contribution absolue à l'information 6 millinats, 3 millinats et 2 millinats. Ces résultats confirment parfaitement ceux que nous avons eu en faisant le choix des modèles.

En considérant le tableau profession x sexe, c'est chez les femmes que les différences sont très significatives (la contribution à l'information est de 308 millinats contre 93 millinats chez les hommes). De même, c'est dans les métiers les plus féminisés (on le verra au chapitre 7) que la contribution à l'information est très forte : les ouvriers qualifiés du travail des étoffes, les métiers de la santé, le personnel qualifié des soins paramédicaux et sociaux, les métiers des soins personnels, personnel non qualifié des techniques comptables ont respectivement 464 millinats, 436 millinats, 452 millinats, 252 millinats et 444 millinats comme contribution absolue de la marge profession.

Quant au tableau profession x région, certaines régions sont très déviantes par rapport à la moyenne : Ile-de-France (54 millinats), Franche Comté (48 millinats) Poitou-Charentes (43 millinats), Provence - Alpes - Côte d'Azur- Corse (38 millinats, contrairement aux régions où les contributions à l'information sont faibles : Nord-Pas-de-Calais (17 millinats), Rhône-Alpes (13 millinats) etc.

En ce qui concerne les professions, les mécanographes, opérateurs, perforateurs (137 millinats), les métiers non qualifiés de l'hôtellerie (94 millinats), les métiers techniques de l'hôtellerie (90 millinats) sont celles où la contribution à l'information est importante et elles s'opposent aux ouvriers qualifiés de la radio-électricité, du papier-carton et autres spécialités de l'alimentation (10 millinats), métiers techniques de la distribution (17 millinats) et autres ouvriers qualifiés (8 millinats) où cette contribution est faible. Cette lecture peut également s'étendre aux autres marges d'ordre 2 du tableau T (21, 23, 6, 2)

Avant de clore ce paragraphe relatif aux résultats statistiques, il serait aussi intéressant de décrire les paramètres estimés des modèles d'association.

6.2.2.3 Effets de paramètres des modèles d'association

Nous étudierons plusieurs cas de figure ici allant d'un tableau où une seule variable est ordinale à celui d'un tableau où les deux variables sont ordinales (cas d'un tableau à deux dimensions). L'examen des ces cas nous suggérera quel modèle ajuste au mieux les données du tableau et en quelque sorte les insuffisances du modèle Log-linéaire de base.

On se propose de construire un tableau de dimension 2 sur la base des résultats obtenus au premier paragraphe du chapitre 7. Autrement dit ce tableau permet de Croiser les cinq groupes de régions avec les cinq groupes de professions, ces groupes ayant été obtenus par le biais d'un indicateur de nature multiplicative.

Au départ, si on applique au tableau le modèle d'équirépartition, la déviance du tableau est de 57, 69. Cette déviance est de 56, 24 dans le cas du modèle d'indépendance, ce qui ne donne qu'une amélioration relative de 2,5 %, Cette amélioration reste très faible.

D'autre part, le modèle saturé n'est rien d'autre qu'une reparamétrisation des données du tableau brut, et il n'a aucun intérêt dans la mesure où l'application du modèle Log-linéaire a surtout pour objet d'ajuster au mieux les données d'un tableau avec une structure plus simple (principe de la parcimonie). Ces deux cas extrêmes suggèrent l'existence d'un modèle intermédiaire qui lisse parfaitement les données du tableau.

Supposons dans un premier temps que l'une des variables du tableau soit ordinale Prenons la variable profession dans ce cas. Le modèle de lissage est le suivant :

$$M_1 : \log t(i_1, i_2) = \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + (s_{i_2} - \bar{s}_2)\alpha'_1(i_1)$$

Les effets de paramètres de ce modèle se trouvent au tableau 16.d. Le modèle M_1 donne de très faibles valeurs de $\alpha'_1(i_1)$ qui sont toutes proches. D'autre part, il n'est pas bon. Sa déviance est de 55, 58 pour 12 degrés de liberté contre 56, 69 pour 16 degrés de liberté en ce qui concerne le modèle d'indépendance.

Si les scores si s_{i_2} ($i_2 = 1, 2, \dots, I_2$) n'étaient pas centrés par rapport à la moyenne, le résultat ne changerait pas pour ce qui est des valeurs de $\alpha'_1(i_1)$ et la déviance.

A l'inverse, si la variable région est ordonnée et la profession nominale, le modèle devient:

$$M_2 : \log t(i_1, i_2) = \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + (s_{i_1} - \bar{s}_1)\alpha'_2(i_2)$$

L'amélioration du modèle d'indépendance est très nette ici. La déviance du modèle est 41, 23, et le modèle exprime environ 27,3 % du nuage des données (cf. tableau 16c). Le modèle M_2 donne les valeurs $\alpha'_2(i_2)$ ci-après:

$$\alpha'_2(1) = 0 ; \alpha'_2(2) = 0.198 ; \alpha'_2(3) = 0.227 ; \alpha'_2(4) = 0.087 ; \alpha'_2(5) = 0.130$$

On constate que ces valeurs ne sont pas sur une échelle ordinale, ce qui signifie l'inexistence d'une association linéaire à travers toute modalité de la variable région. Ce modèle peut être combiné au modèle M_1 et donne M_3 :

$$M_3: \log t(i_1, i_2) = \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + (s_{i_1} - \bar{s}_1)\alpha'_2(i_2) + (s_{i_2} - \bar{s}_2)\alpha'_1(i_1)$$

La déviance du modèle M_3 n'est que très légèrement différente de celle du modèle M_2 , ce qui reflète parfaitement la quasi-nullité du terme $(s_{i_1} - \bar{s}_1)\alpha'_2(i_2)$ dans l'amélioration du modèle d'indépendance. Il a pour déviance 40,77 contre 41,23 pour le modèle M_2 .

Si toutes les deux variables du tableau sont ordinales, et que l'on suppose connus a priori les scores alors le modèle d'association uniforme est tel que :

$$M_4: \log t(i_1, i_2) = \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \gamma(s_{i_1} - \bar{s}_1)(s_{i_2} - \bar{s}_2)$$

L'estimation du modèle M_4 donne les effets de paramètres indiqués au tableau 16. Il est encore mauvais par rapport au modèle d'indépendance puisque le terme supplémentaire γ n'apporte aucune information substantielle. La valeur $\gamma = 0,010$

On peut donc constater que les modèles M_1 , M_2 , M_3 et M_4 sont de mauvaise qualité, même si le modèle M_2 améliore relativement mieux le modèle d'indépendance. Peut-être que le modèle log-multiplicatif est-il susceptible d'apporter des améliorations sensibles.

En rappelant le modèle log-multiplicatif, il est tel que :

$$M_5: \log t(i_1, i_2) = \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \gamma_i s_{i_2}$$

les scores s_{i_1} et s_{i_2} étant inconnus.

L'estimation du modèle nous a permis d'aboutir aux ensembles de scores ci-après :

-0.722	0.452	-0.278	0.427	0.121	Scores lignes centrés et normés
-0.518	0.697	0.139	-0.456	0.118	Scores colonnes centrés et normés

La valeur de γ est de 0,878.

La somme de chacun des ensembles de scores est nulle ; la somme de leur carré est égale à l'unité d'où la validation des contraintes de normalisation.

La convergence du modèle a été obtenue au bout de 4 itérations et sa déviance est 7.23. Il est nettement meilleur que les modèles M_1 , M_2 , M_3 , M_4 et il exprime environ 87,20 % de l'inertie totale des données. C'est le modèle qui ajuste assez bien les données du tableau en Annexe.

Tableau 17 : Statistiques du G2 et du Pearson

Nombre d'itérations	Déviance	Statistique de Pearson
1	25.05	24.83
2	7.969	8.196
3	7.295	7.542
4	7.285	7.528

Tableau 18 : Mesure d'association par le modèle log-multiplicatif

0.328	-0.442	-0.088	0.289	-0.087
-0.205	0.276	0.055	-0.181	0.054
0.126	-0.170	-0.034	0.091	-0.033
-0.194	0.258	0.052	-0.171	0.052
-0.055	0.074	0.015	-0.048	0.015

Tableau 19 : estimation des "scores -lignes" et "scores -colonnes" centrés

-1.615	1.011	-0.621	0.954	0.271	Scores lignes centrés
-1.157	1.559	0.311	-1.020	0.308	Scores colonnes centrés

Le modèle M_5 indique qu'il y a une proximité (structurelle) forte entre les régions du groupe II et celles du groupe IV (cf. scores lignes). D'ailleurs, la distance des scores associés

à ces deux groupes est de 0,025 seulement. De même, il y a une forte proximité entre les professions du groupe III et les professions du groupe V (l'écart entre les scores est de 0,0014)

En examinant les tableaux 18 et 19 on voit de façon plus pertinente l'un des points forts du modèle log-multiplicatif. Du fait de son approche globalisante le modèle Log-linéaire de base n'est pas capable de ressortir efficacement les associations locales que nous constatons dans certaines régions du tableau 18, ce que reflète parfaitement le modèle log-multiplicatif. De plus, les proximités structurelles esquissées par ce dernier modèle ne sont pas assez visibles avec le modèle Log-linéaire de base.

Toute permutation des "scores lignes" et/ou des "scores colonnes" laisse le ratio de la distance relative invariant. Il est alors possible de réordonner les groupes de régions et les groupes de professions de telle sorte que les modalités de chacun des groupes se situent sur une échelle ordinale (nous rappelons que l'affectation de chaque groupe de professions à chaque groupe de régions a été arbitraire). Autrement dit, le modèle log-multiplicatif peut suggérer quel groupe de professions associer à chaque groupe de régions, ce qui illustre un autre aspect de sa richesse.

Les modèles que nous venons d'explorer s'appliquent également pour le lissage d'un tableau de dimension supérieure à deux.

Soient les groupes de régions et les groupes de professions constitués précédemment. La prise en compte de l'âge et du sexe nous amène à former le tableau T (5, 5, 6, 2) de dimension 4 (cf. annexe A.6) qui a 300 cellules.

Si la variable âge est supposée ordinale. on peut avoir le modèle suivant pour des scores arbitraires associés aux modalités de l'âge :

M_6

$$\log t(i_1, i_2) = \alpha_0 + \alpha_1(i_1) + \alpha_2(i_2) + \alpha_3(i_3) + \alpha_4(i_4) + (s_{i_1} - \bar{s}_3)\alpha'_2(i_2) + (s_{i_1} - \bar{s}_3)\alpha'_1(i_1) + (s_{i_1} - \bar{s}_3)\alpha'_1(i_1)$$

Ce modèle indique l'existence des différentes associations conditionnelles, entre l'âge et chacune des autres variables. Les paramètres estimés de ce modèle se trouvent au tableau 20 et nous les présentons à titre illustratif, car ces paramètres étant non centrés comme ceux des modèles M_1 , M_2 , M_3 , M_4 , nous ne pouvons leur donner de signification.

Outre, la présentation des résultats statistiques, nous allons entrer dans la phase analytique de ces résultats. L'analyse portera sur les effets de paramètres centres dont la significativité a été déjà mentionnée plus haut. Le paragraphe qui suit sera axé sur l'analyse des effets de premier ordre.

6.3. ANALYSE DES EFFETS SIMPLES

L'application du modèle Log-linéaire dans son approche "extensive" a permis d'estimer les effets de paramètres liés aux marges unidimensionnelles du tableau d'étude défini au paragraphe 6.1. Ces effets de paramètres sont des effets de niveau ou de proportionnalité et reflètent l'importance relative de chaque modalité d'une variable dans la marge d'ordre zéro (effectif total de l'échantillon). Nous allons recenser ces différents effets, recherchant à mettre l'accent sur leurs profils éventuels, tant en niveau qu'en évolution. Chaque fois qu'on le pourra, des éléments d'explication seront apportés aux évolutions significatives.

Il ressort des résultats, que la variable la plus discriminante est la variable profession. Viennent ensuite dans l'ordre, les variables région, sexe et âge. Nous retiendrons cet ordre dans la présentation des effets, toutefois en analysant les effets sexe et âge ensemble, dans la mesure où ces dernières variables ont moins de modalités que les deux premières. Ainsi, ferons nous successivement les analyses suivantes :

- l'analyse de l'effet profession,
- l'analyse de l'effet région,
- l'analyse de l'effet sexe et de l'effet âge.

6.3. 1 Analyse de l'effet profession

La variable profession est celle qui présente d'importantes différences tant en niveau qu'en évolution. Autrement dit, il existe de très fortes disparités entre les métiers, ce à travers le temps.

A titre d'exemple, les ouvriers spécialisés et manœuvre ont un poids relatif de 22 % environ contre 0.6 % pour les métiers techniques de l'hôtellerie en 1976. Ces éléments nous invitent à ébaucher des profils en niveau et en évolution des métiers afin d'apprécier dans un

premier temps, les métiers qui ont un poids relativement semblable et dans un deuxième temps ceux qui ont des évolutions assez ou peu significatives.

6.3 .1 .1 Profils différenciés des professions en niveau

En niveau , on peut distinguer trois profils caractérisant les métiers (cf. graphique 6b) .

- les métiers à très forte représentativité par rapport à la moyenne et qui comprennent les ouvriers spécialisés et manœuvres (22 % , les autres ouvriers qualifiés (19,7 % ce dernier métier étant assez hétérogène puisqu'il regroupe le reste des ouvriers qualifiés non signalés dans la nomenclature à 23 métiers retenue ici (cf. paragraphe 6.1). Il n'est donc pas surprenant qu'on le retrouve dans ce groupe.

- les métiers à représentativité moyenne, parmi lesquels les métiers qualifiés de la vente et hôtellerie (8,4 techniciens et dessinateurs (6,2 %), employés non qualifiés (5,6 %), etc.

- les métiers à faible représentativité par rapport à la moyenne notamment ouvriers qualifiés du travail des étoffes (1,5 %), employés administratifs qualifiés (1.2%) etc. Ce groupe renferme toutefois des métiers très sous-représentés tels que ouvriers qualifiés verre et photographie (0,74) et métiers techniques de l'hôtellerie (0,6%).

Cette typologie reste quasi-stable sur la période 1976-1981, ce qui ne met aucunement en cause les évolutions contrastées qui sont observées à l'intérieur des métiers.

6.3.1.2. Profils différenciés des professions en évolution

Il se dégage au vu des résultats, deux profils d'évolution significative : les professions dont les effectifs augmentent régulièrement sur la période 1976- 1981 et celles dont les effectifs baissent régulièrement sur la même période. Comme on le verra, il s'agira d'une opposition entre métiers du tertiaire à haut niveau de qualification et métiers du secondaire à bas niveau de qualification.

Tous les métiers dont les effectifs augmentent sur la période 1976-1981 sont les métiers qualifiés du tertiaire à l'exception des métiers de la santé qui ont un niveau de qualification assez bas.

La progression des effectifs est spectaculaire dans les métiers de la santé chez le personnel diplômé des services paramédicaux et sociaux, les professeurs, professions littéraires et de l'information, les cadres moyens ; elle est moins, soutenue mais assez régulière chez les cadres administratifs supérieurs (cf. graphique 7). Ces évolutions, bien que similaires sont liées à des facteurs totalement différents.

Les métiers de la santé et personnel diplômé des services paramédicaux et sociaux sont bien entendu les métiers de la branche santé, dans laquelle le rôle des pouvoirs publics est prépondérant comme régulateur de l'offre et la demande des soins. En effet, le développement de la production des services de la santé ne dépend pas seulement des ménages, mais également de l'action des pouvoirs publics qui "garantissent le financement socialisé de la demande des soins".

Jusqu'en 1979, où intervient la politique de rationalisation de l'offre des soins se traduisant par une modération de la progression des dépenses de la santé , la demande des soins est restée croissante et le tassement des effectifs des métiers de la santé à partir de cette période semble refléter les effets de cette politique.

La progression des effectifs chez les cadres moyens et les cadres administratifs supérieurs est à lier au développement du tertiaire voire son extension à l'industrie. Cette extension fait par exemple apparaître de nouvelles fonctions dans des entreprises (administration, gestion, informatique, distribution etc.) confiées aux cadres et la conséquence de ces nouvelles tâches est la croissance des effectifs des cadres.

D'autre part, dans le contexte de la crise actuelle, certaines entreprises ont développé une politique de rationalisation de la main-d'œuvre qui les a poussées plutôt à orienter leurs embauches vers les cadres.

Les professeurs, professions littéraires et de l'information voient leurs effectifs augmenter sur la période du fait, semble-t-il de la croissance des enseignants qui ne tient qu'à la prolongation de la scolarité. Encore faudra-t-il nuancer ici puisque la catégorie comprend également les professions de l'information qui n'ont aucun lien avec l'enseignement.

On peut rattacher à ce groupe de métiers à effectifs croissants les comptables et techniciens des banques dont les effectifs après avoir été stabilisés entre 1976 et 1978 augmentent lentement, mais de façon régulière jusqu'en 1981.

A ce groupe s'opposent des métiers qui enregistrent plutôt une baisse des effectifs entre 1976 et 1981 et qui sont les métiers du secondaire à bas niveau de qualification, notamment les ouvriers spécialisés et manœuvres, les ouvriers qualifiés du travail des étoffes et les ouvriers qualifiés du textile, cuirs et peaux, chimie (cf. graphique 7a). On y trouve également un métier du tertiaire à savoir les métiers des soins personnels.

La crise a frappé avec acuité les activités industrielles et la baisse de l'emploi est surtout sensible dans les catégories ouvrières. Ceci justifierait la baisse des effectifs assez prononcée chez les ouvriers spécialisés et manœuvres.

On peut voir également dans la baisse des effectifs des ouvriers spécialisés et manœuvres voire des ouvriers en général (qualifiés et non qualifiés) une nécessité d'adaptation aux contraintes de l'environnement international ou de restructuration de l'appareil productif. En effet, la recherche de la productivité et de la compétitivité engendre dans la situation de crise actuelle une transformation du processus de production se traduisant par une automatisation plus poussée, le résultat en termes de suppressions d'emplois ouvriers étant évident.

La crise aiguë que connaît le textile-habillement n'est pas aussi neutre dans la baisse des effectifs notamment dans les métiers d'ouvriers qualifiés du travail des étoffes et ouvriers qualifiés du textile, cuirs et peaux, chimie. Elle n'a fait qu'accentuer un mouvement amorcé depuis les années soixante du fait de la baisse de la demande des ménages à laquelle le textile-habillement est très sensible. En 20 ans, les effectifs y ont baissé de 370 000, soit à un rythme de 1,8% l'an entre 1960 et 1975 et 4% l'an entre 1975 et 1980.

Les autres métiers ont des évolutions soit stables (métiers techniques de la distribution, métiers qualifiés de la vente et hôtellerie etc.) soit non significatives (cf. Annexe A.3).

L'analyse de l'effet profession nous amène à envisager également celle de l'effet région.

6.3.2. Analyse de l'effet région

La région a un pouvoir de différenciation moins prononcé que la profession et ceci sera surtout visible en analysant l'évolution des effectifs régionaux dans notre échantillon. Ainsi, la région qui connaît un taux de chômage élevé n'est pas forcément celle où les effectifs salariés baissent. C'est, qu'au niveau des régions, il y a une relative étanchéité du marché de travail où se superposent plusieurs facteurs et qui jouent simultanément. Comme dans les professions, nous analyserons ici les profils différenciés des régions en niveau et en évolution

6.3.2.1 Profils différenciés des régions

On peut concevoir quatre types de régions (cf. graphique 6a)

- les régions à très forte représentativité par rapport à la moyenne. C'est le cas de l'Île-de-France (24,9%) qui appartient seule à ce groupe.

- les régions qui ont une forte représentativité par rapport à la moyenne Rhône-Alpes (9,3%), Nord Pas de Calais (7.2%), Provence Alpes-Cote d'Azur-Corse (6,6%).

- les régions à représentativité moyenne : Lorraine (4,4%), L'Alsace (3,1%), la Bretagne (3,9 %), etc.

les régions à faible représentativité par rapport à la moyenne, l'Auvergne Basse Normandie (2,2%), la Franche comté et le Limousin, ces deux régions étant les plus sous-représentées par rapport à la moyenne avec des poids respectifs de 1.9% et 1,1% d'effectifs salariés. Ce groupe comprend bien entendu d'autre régions que nous n'avons pas mentionné.

Cette typologie est en correspondance parfaite avec le poids démographique des régions et l'information qui en découle n'est donc pas substantielle. Il semble dès lors qu'on ne peut tirer profit que d'une analyse en évolution.

6.3.2.2. Profils différenciés des régions un évolution

L'évolution des effectifs salariés régionaux n'apparaît pas de façon tranchée. En effet , seul le tiers des régions a une évolution modestement significative, les autres se caractérisant par la quasi-stabilité de leurs effectifs sur la période 1976-1981.

Le groupe de région à évolution significative comprend : Ile-de-France, la Lorraine, l'Alsace, la Picardie, le Languedoc-Roussillon et à un degré moindre la Franche comte et le Rhône-Alpes.

Contrairement aux professions où il y avait un clivage tertiaire-secondaire dans les évolutions, ici l'activité dominante semble n'avoir qu'une influence limitée sur l'évolution des effectifs.

L'évolution la plus nette est à mettre à l'actif de la région Languedoc-Roussillon (cf. graphique 8) qui voit ses effectifs salariés augmenter sur la période 1976-1981, malgré un tassement de ces effectifs à partir de 1979. Il s'agit bien entendu d'une région tertiariée et où l'emploi a été très soutenu sur la période 1975-1981. En effet, la croissance de l'emploi global est la plus forte du territoire (10 %) sur cette période.

D'ailleurs; la région Languedoc-Roussillon se singularise par son attractivité depuis 1975 - l'indice d'attractivité est la plus forte, sur le plan national (1.07%) - le flux migratoire dont la région bénéficie (ce flux est l'un des plus élevés) notamment celui des retraites (phénomène de "retour au pays ") joue un rôle moteur sur la demande des services, laquelle a des répercussions évidentes en matière de création d'emplois.

Les effectifs salariés progressent également en Rhône-Alpes jusqu'en 1978 avant de se stabiliser. Cette évolution est assez surprenante lorsque l'on sait que sur la période 1974-1984, elle a perdu 100.000 emplois industriels nonobstant de nombreuses reconversions et le rôle moteur qu'y joue désormais le tertiaire du fait de son urbanisation. D'ailleurs, l'emploi global a décliné en Rhône-Alpes au rythme de 0,03% l'an sur la période 1975-1981.

A l'inverse, les effectifs salariés régressent en Ile-de-France et en Lorraine. Si cette évolution semble concordante avec le déclin de l'industrie traditionnelle et qui se traduit par des suppressions d'emplois en Lorraine, il n'en est pas de même pour l'Ile-de-France qui enregistre une croissance de près de 1% de l'emploi global entre 1975-1981.

Ces différentes évolutions, qui ne reflètent pas toujours l'état de l'activité d'une région nous convient à rechercher ailleurs des éléments qui jouent un rôle moteur dans la dynamique de l'emploi régional, puisque le Nord Pas-de Calais, par exemple longtemps touché par la crise et en proie au problème de reconversion voit ses effectifs se stabiliser entre 1976-1981. Même s'il n'existe pas de schéma de causalité évolution de l'emploi - évolution du chômage,

on peut se poser des questions sur cette relative stabilité. Autant d'éléments qui indiquent la structure complexe des marchés du travail régionaux et dont les facteurs qui les façonnent apparaîtront en prospectant les effets de structure (cf. chap. VII). Avant d'en arriver, il est nécessaire d'analyser les effets âge et sexe.

6.3-3 Analyse des effets âge et sexe

Nous examinerons tout d'abord l'effet âge avant d'en faire autant pour l'effet sexe

6.3.3.1 Analyse de l'effet âge

L'analyse de l'effet âge ne nous permet pas de trouver en niveau des profils quasi-stables sur la période 1976-1981, même s'il apparaît que la classe d'âge 25-32 ans est la plus représentée sur la période et les classes d'âge 48-55 ans et 55 ans et plus. les moins représentées (cf., graphique 6c). Les modalités sont relativement assez homogènes ici par rapport à celles des variables profession et région : la classe d'âge 25-32 ans qui est la plus représentée a un poids relatif de 21,7% contre 11,6% pour la classe 55 ans et plus qui est la moins représentée

En évolution, les effectifs des classes d'âge 32-40 ans, 40-48 ans, 48-55 ans, 55 ans et plus se singularisent par leur quasi-stabilité. Les deux autres classes ont des évolutions totalement opposées (cf. graphique 9).

La classe d'âge des moins 25 ans enregistre une baisse continue des effectifs et les raisons de cette décroissance seraient d'ordre socio-économique.

La crise a entraîné le freinage de l'embauche des jeunes de moins de 25 ans et. l'une des raisons évoquées est qu'ils n'ont aucune expérience professionnelle. C'est ce qui justifie en partie la "vulnérabilité" des jeunes au chômage (24% de jeunes actifs et 31% de jeunes actives sont au chômage en 1984) et la baisse de leur effectifs salariés. Cette baisse est de l'ordre de 19% jusqu'en 1981 où il y a une inflexion probablement due aux effets des contrats solidarité-préretraite et des mesures prises dans le cadre des pactes pour l'emploi.

La prolongation de la scolarité est aussi un élément à lier à la décroissance des effectifs des jeunes. À cela, on peut ajouter les raisons induites par l'arrivée à l'âge adulte des jeunes

recrutés dans la période précédente et une analyse longitudinale nous aurait sûrement permis de saisir de façon plus pertinente ce phénomène.

A l'inverse de ce mouvement, les effectifs ne cessent de croître à l'intérieur de la classe d'âge 25-32 ans, ce qui reflète à juste titre l'entrée massive des femmes sur le marché de travail. Elles sont en général dans la tranche d'âge 25-54 ans où le mouvement de féminisation a été amorcé depuis. Il y a donc une modification structurelle de l'emploi qu'engendrent de profondes mutations dans le processus de production, la conséquence étant le recours à une main-d'œuvre de réserves laquelle se situe en général dans la tranche d'âge 25-32 ans (25-54 plus globalement) d'où la croissance des effectifs enregistrée ici.

Un autre élément qui joue en faveur de cette croissance est l'arrivée sur le marché du travail, dans la période 1974-1981 de nombreuses générations aux âges de 20-35 ans, consécutive à la forte augmentation de la natalité de l'après-guerre. Nous quitterons le cadre de cette analyse pour apprécier celle de l'effet sexe.

6.3-3.2 Analyse de l'effet sexe

Quelle que soit l'année, on constate une très forte représentativité des hommes qui s'oppose à une très faible représentativité des femmes (64,6% d'hommes contre 33,4% de femmes, cf. graphique 6d).

Si la part des effectifs masculins reste relativement stable entre 1976 et 1981, celle des femmes croît légèrement (cf. graphique 10), ceci est le résultat du phénomène de forte féminisation amorcée depuis les années soixante et dont nous avons déjà ébauché quelques éléments d'explication dans le précédent paragraphe, notamment la transformation de l'appareil productif et qui permet de faire appel à la main-d'œuvre féminine. D'autre part, avec le développement du tertiaire, c'est autant d'emplois requérant une main-d'œuvre aux caractéristiques généralement féminines qui sont créées.

D'autres raisons d'ordre sociologique peuvent être avancées comme la baisse de la fécondité. Celle-ci a pour effet de rendre plus continue l'activité des femmes, puisque le taux d'activité féminin décroît avec le nombre d'enfants.

CONCLUSION

La définition du tableau des données nous a conduit à construire pour chaque année sur la période 1976-1981 un tableau de contingence multidimensionnel $T(21, 23, 6, 2)$, auquel on applique la méthodologie Log-linéaire pour en décrypter les structures latentes. Certes, on n'a pu mettre en relief pour l'instant que l'aspect surtout exploratoire des résultats statistiques.

Le meilleur modèle, dans l'approche du modèle Log-linéaire de base reste un outil de validation des liaisons globales dans un tableau. Il ne peut déceler les liaisons locales à l'intérieur du tableau, ce qui reste une des richesses du modèle Log-multiplicatif, qui s'apparente bien à la méthode d'analyse factorielle des correspondances, même si le fondement des deux approches est différent.

Le modèle Log-multiplicatif a été utilisé pour lisser un tableau de dimension 2. On aurait pu en faire autant pour les tableaux de dimension supérieure.

De même l'analyse des effets simples a été conduite ici et on a pu voir des différences très nettes entre les modalités de la variable profession. Cependant, bien que l'analyse des effets ait un intérêt, l'information qu'on en tire reste encore modeste du fait même de son caractère globalisant à ce niveau. D'autre part, l'un des objets de la méthodologie étudiée est la prospection de l'information à l'intérieur d'un tableau ou d'un sous-tableau. Autrement dit, comment caractériser par exemple les caractéristiques régionales des salariés à travers les métiers ? Les femmes investissent-elles indifféremment les métiers ? La réponse à de telles questions passe indubitablement par l'analyse des effets de structure que nous explorerons au chapitre suivant.

ANNEXE

CYCLE	SCALED DEVIANCE	DF	ESTIMATE	S. E.	PARAMETER
4	532.6	299			
1			1.733	.803E-01	ZGM
0			ZERO	ALIASED	A(1)
2			.3239	.7796E-01	A(2)
3			-1.677	.1267	A(3)
4			-.5363	.1117	A(4)
5			-.6756	.1763	A(5)
0			ZERO	ALIASED	B(1)
6			.2911	.8562E-01	B(2)
7			-.2615	.1103	B(3)
8			-.8204	.1429	B(4)
9			.4314	.1436	B(5)
10			.1011E-01	.1231E-01	X
SCALE PARAMETER TAKEN AS					1.000

Tableau 16.a

CYCLE	SCALED DEVIANCE	DF	ESTIMATE	S. E.	PARAMETER
4	532.6	299			
1			1.722	.760E-01	ZGM
0			ZERO	ALIASED	A(1)
2			.3592	.6307E-01	A(2)
3			-1.016	.1017	A(3)
4			-.4453	.1374E-01	A(4)
5			-.5541	.1670E-01	A(5)
0			ZERO	ALIASED	B(1)
6			.3215	.702E-01	B(2)
7			-.2009	.1101E-01	B(3)
8			-.7294	.1030	B(4)
9			.5528	.799E-01	B(5)
10			.1011E-01	.1231E-01	X
SCALE PARAMETER TAKEN AS					1.000

Tableau 16.b

CYCLE	SCALED DEVIANCE	DF	ESTIMATE	S. E.	PARAMETER
4	517.6	297			
1			1.924	.9236E-01	ZGM
0			ZERO	ALIASED	A(1)
2			.2325	.7874E-01	A(2)
3			-1.275	.1299	A(3)
4			-.8422	.1504	A(4)
5			-1.094	.1916	A(5)
0			ZERO	ALIASED	B(1)
6			-.1829	.1683	B(2)
7			-.7800	.1936	B(3)
8			-.9539	.2227	B(4)
9			.2156	.1592	B(5)
10			ZERO	ALIASED	X1
0			ZERO	ALIASED	B(1), X1
11			.1930	.5941E-01	B(2), X1
12			.2237	.6661E-01	B(3), X1
13			.8739E-01	.7776E-01	B(4), X1
14			.1302	.5747E-01	B(5), X1
SCALE PARAMETER TAKEN AS					1.000

Tableau 16.c

CYCLE	SCALED DEVIANCE	DF	ESTIMATE	S. E.	PARAMETER
4	531.2	297			
1			1.798	.9577E-01	ZGM
0			ZERO	ALIASED	A(1)
2			.2354	.1527	A(2)
3			-1.191	.2305	A(3)
4			-.5677	.1082	A(4)
5			-.6987	.1953	A(5)
0			ZERO	ALIASED	B(1)
6			.2845	.8474E-01	B(2)
7			-.2750	.1076	B(3)
8			-.8409	.1384	B(4)
9			.4037	.1410	B(5)
10			ZERO	ALIASED	X2
0			ZERO	ALIASED	A(1), X2
11			.4059E-01	.4463E-01	A(2), X2
12			.5737E-01	.6631E-01	A(3), X2
13			.4082E-01	.5473E-01	A(4), X2
14			.4037E-01	.5653E-01	A(5), X2
SCALE PARAMETER TAKEN AS					1.000

Tableau 16.d

SCALED
CYCLE 4 424.80 DF 284

CYCLE	ESTIMATE	S.E.	PARAMETER
1	1.750	1.657	X(1)
0	ZERO	ALIASED	A(1)
2	.6129	1.521	A(2)
3	-1.171	2.344	A(3)
4	-1.2463	1.961	A(4)

5	-1.4634	1.943	A(5)
0	ZERO	ALIASED	B(1)
6	.0804	1.782	B(2)
7	-1.1468	2.000	B(3)
8	-1.6192	2.431	B(4)
9	.7591	1.745	B(5)
0	ZERO	ALIASED	C(1)
10	.3302	1.621E-01	C(2)
11	.2264	1.292	C(3)
12	.3020	1.642	C(4)
13	.2899	2.044	C(5)
14	.2185	2.443	C(6)
0	ZERO	ALIASED	D(1)
15	.3944	1.203	D(2)
16	ZERO	ALIASED	X 1
0	ZERO	ALIASED	A(1)
17	-1.7854E-01	4.193E-01	A(2)
18	.4616E-01	6.180E-01	A(3)
19	-1.6042E-01	5.144E-01	A(4)
20	-1.2628E-01	5.297E-01	A(5)
0	ZERO	ALIASED	B(1)
21	.1769	4.944E-01	B(2)
22	-1.1847E-01	5.819E-01	B(3)

23	-1.3649E-01	6.513E-01	B(4)
24	-1.6895E-01	4.662E-01	B(5)
0	ZERO	ALIASED	D(1)
25	-1.6781E-01	3.873E-01	D(2)

SCALE PARAMETER TAKEN AS 1.0000

Tableau 20

SCALED
CYCLE 4 517.1 DF 284

CYCLE	ESTIMATE	S.E.	PARAMETER
1	1.949	9.992E-01	X(1)
0	ZERO	ALIASED	A(1)
2	.1604	1.465	A(2)
3	-1.362	2.307	A(3)
4	-1.8464	2.015	A(4)
5	-1.102	2.006	A(5)
0	ZERO	ALIASED	B(1)
6	-1.1915	1.683	B(2)
7	-1.8055	1.976	B(3)
8	-1.9946	2.323	B(4)
9	.1530	1.842	B(5)
10	ZERO	ALIASED	X 1
11	ZERO	ALIASED	X 2

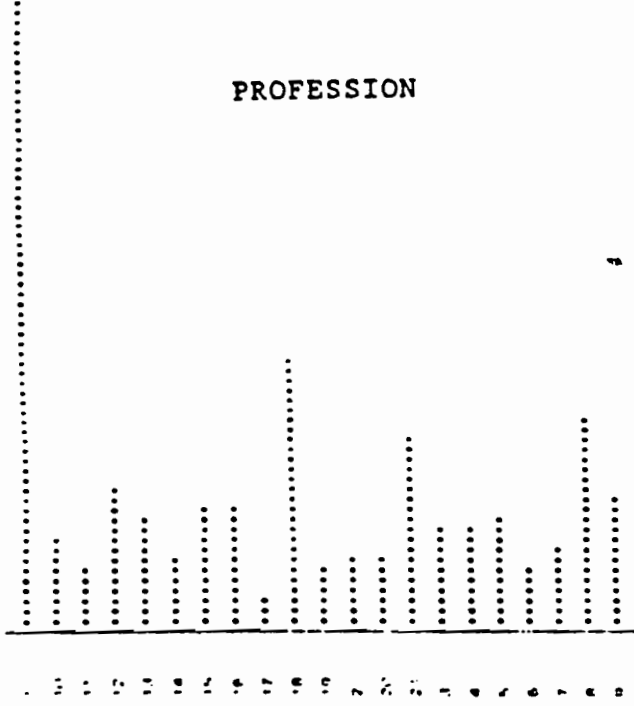
0	ZERO	ALIASED	B(1)
12	.1972	5.979E-01	B(2)
13	.2254	6.808E-01	B(3)
14	.9092E-01	3.822E-01	B(4)
15	.1361	6.441E-01	B(5)
0	ZERO	ALIASED	A(1)
16	.2274E-01	3.955E-01	A(2)
17	.2683E-01	6.424E-01	A(3)
18	-1.5121E-03	5.719E-01	A(4)
19	ZERO	ALIASED	A(5)

SCALE PARAMETER TAKEN AS 1.0000

Tableau 21

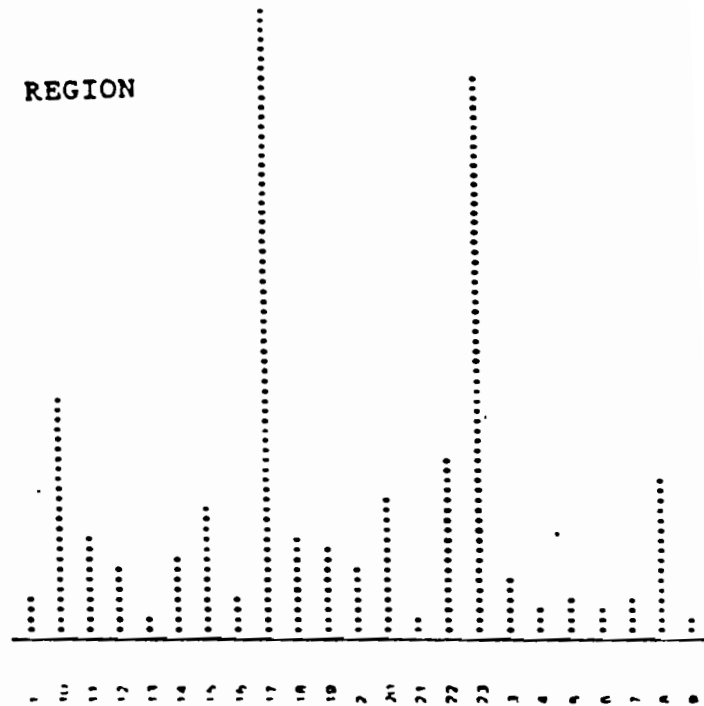
HISTOGRAMME DES EFFECTIFS SALARIES SUIVANT LES VARIABLES

PROFESSION



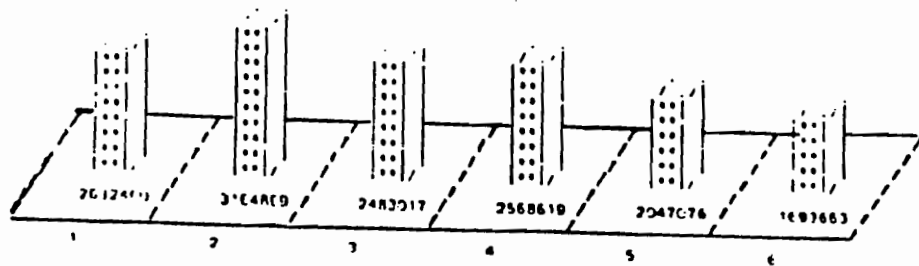
Graphique 6.a

REGION



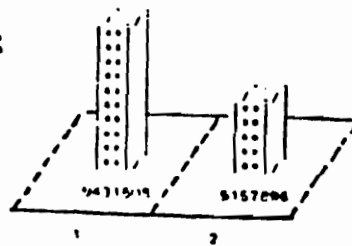
Graphique 6.b

AGE



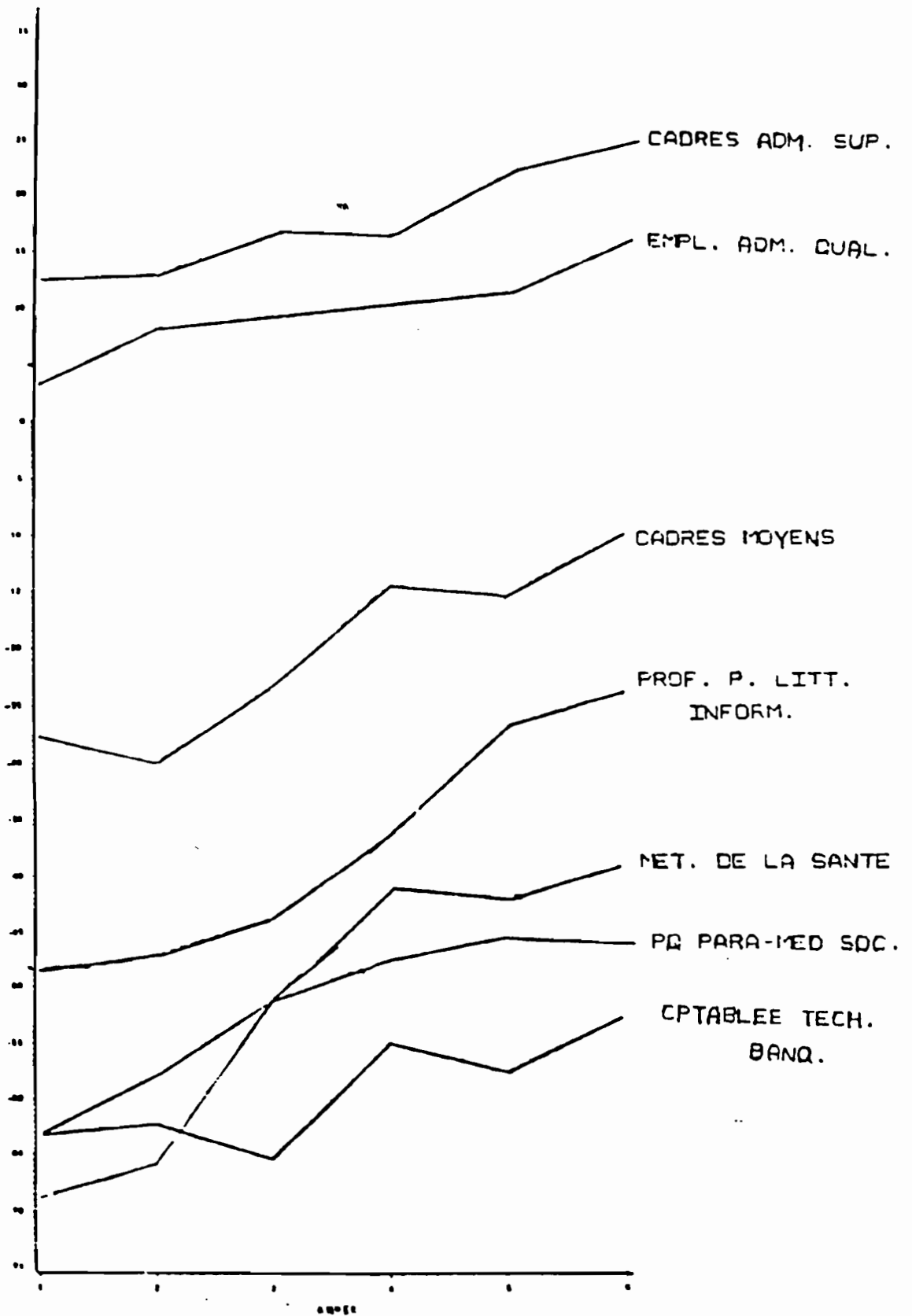
Graphique 6.c

SEXE



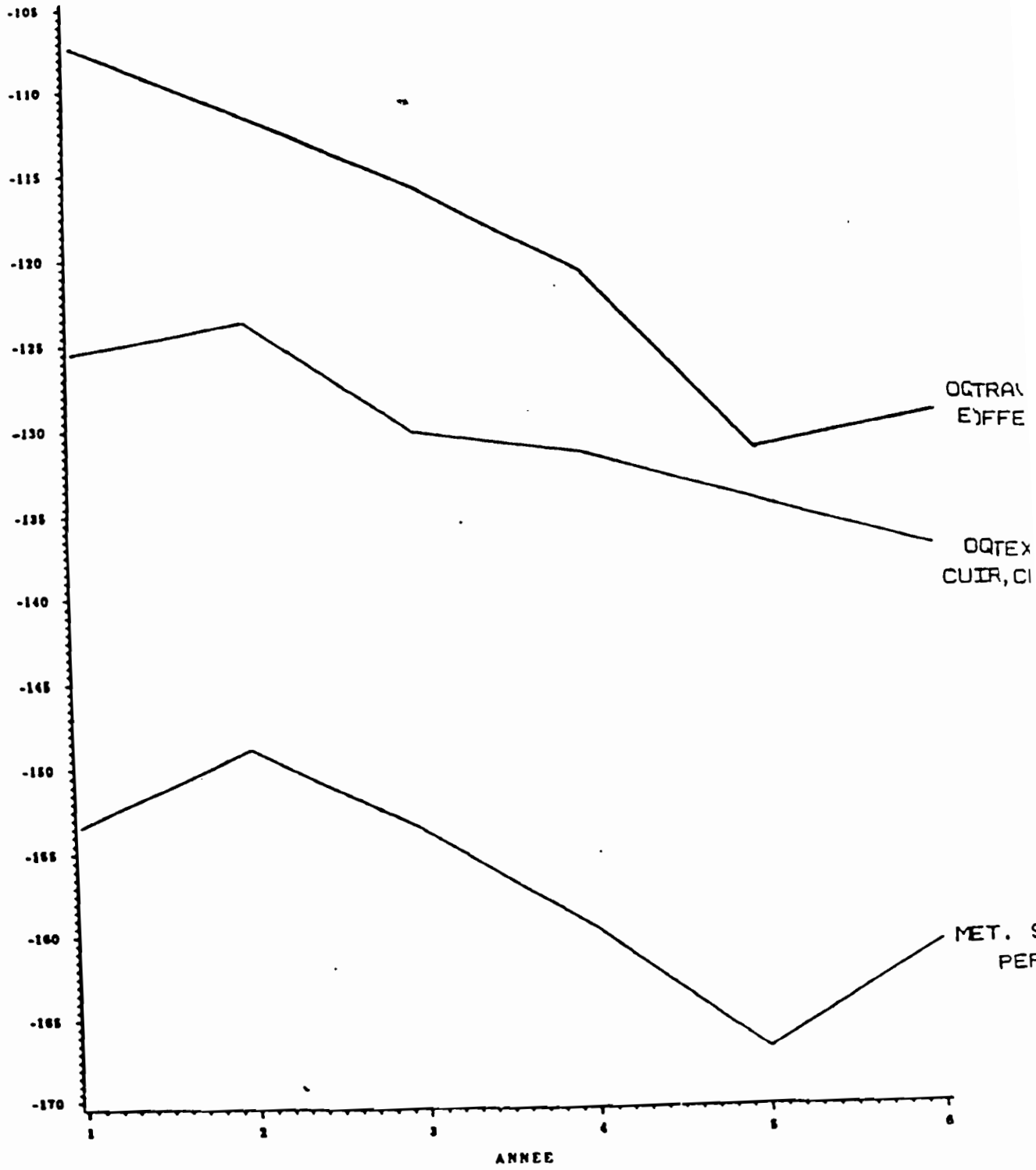
Graphique 6.d

EVOLUTION EFFET PROFESSION (1976-1981)



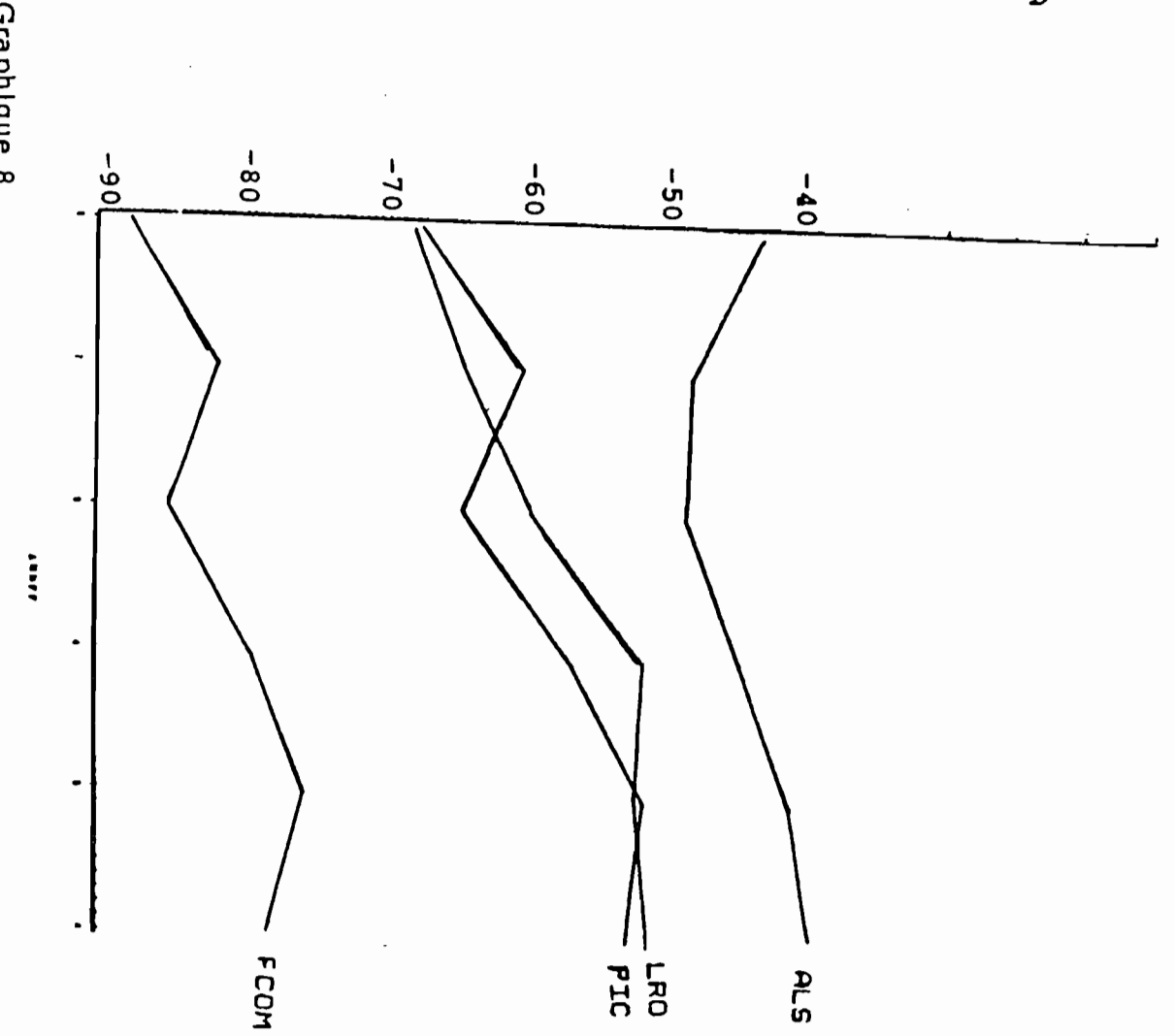
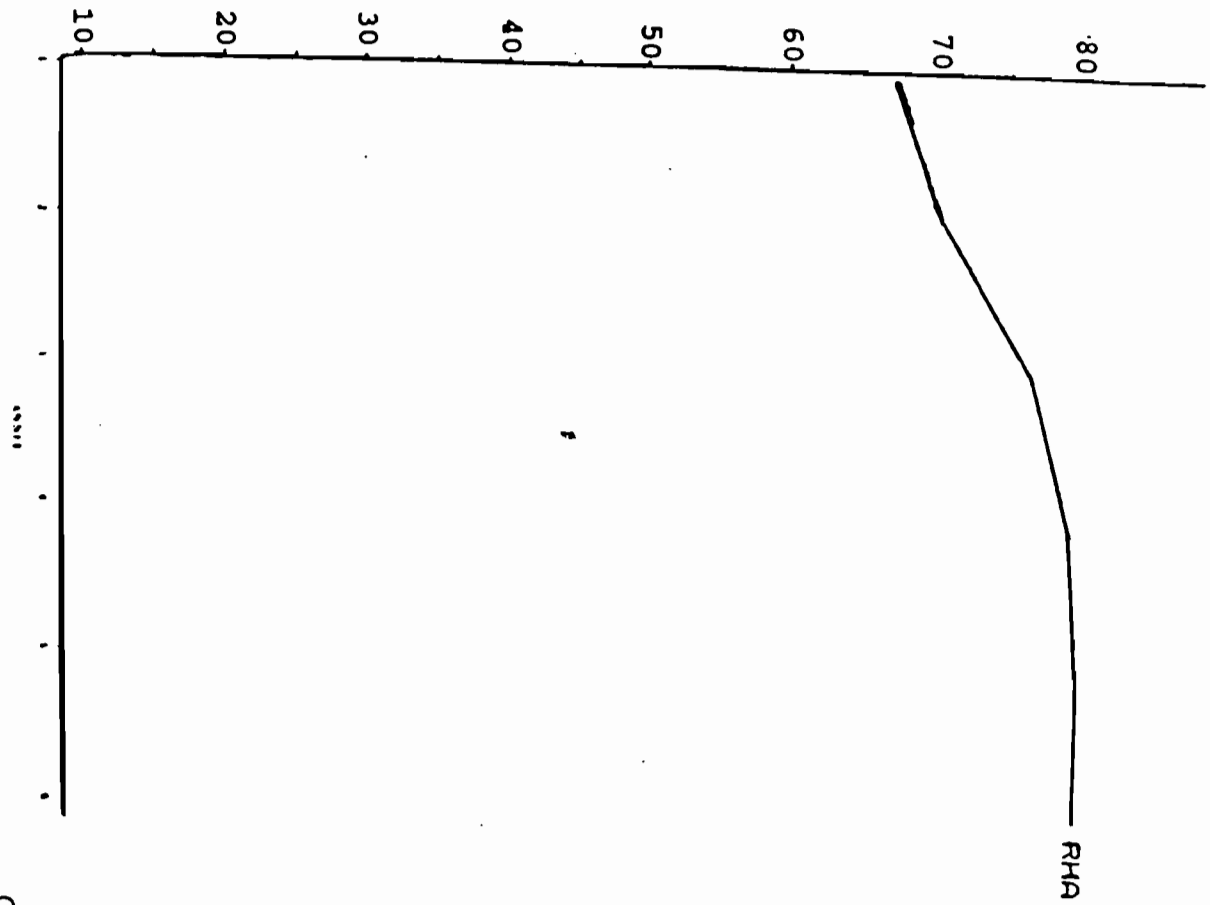
Graphique 7

EVOLUTION EFFET PROFESSION (1976-1981)



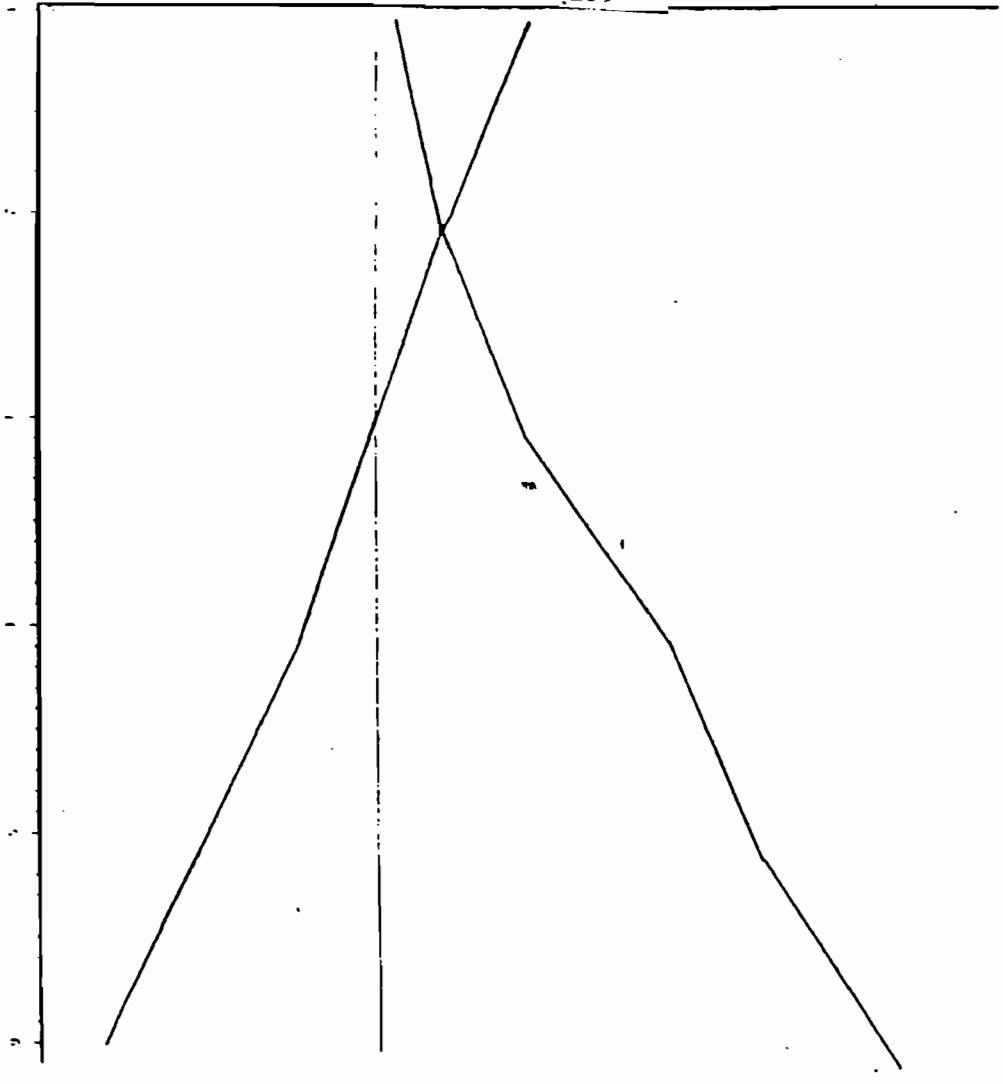
Graphique 7a

EVOLUTION EFFET REGION (1976-1981)



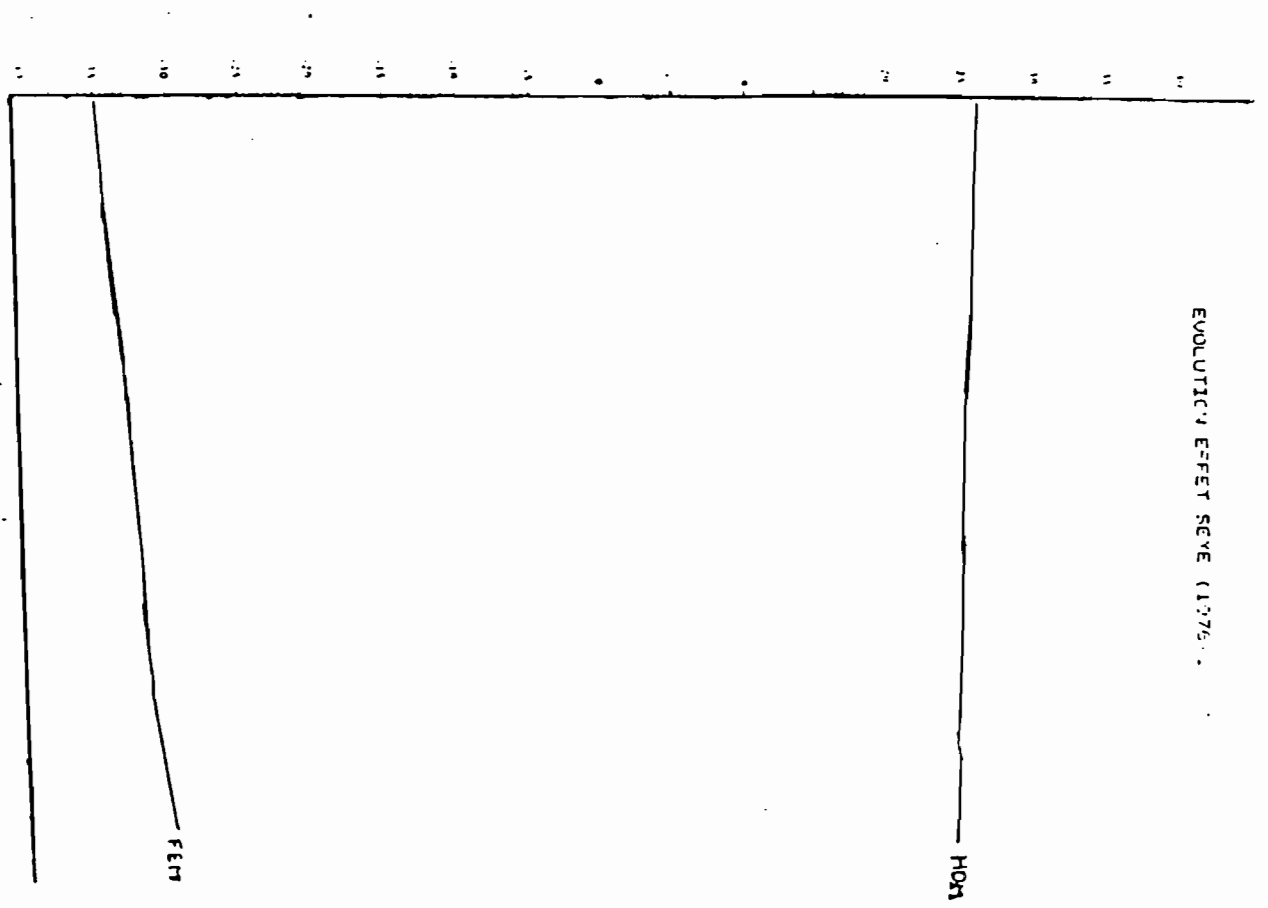
Graphique 8

EVOLUTION DE LA PÉRIODE AGEE



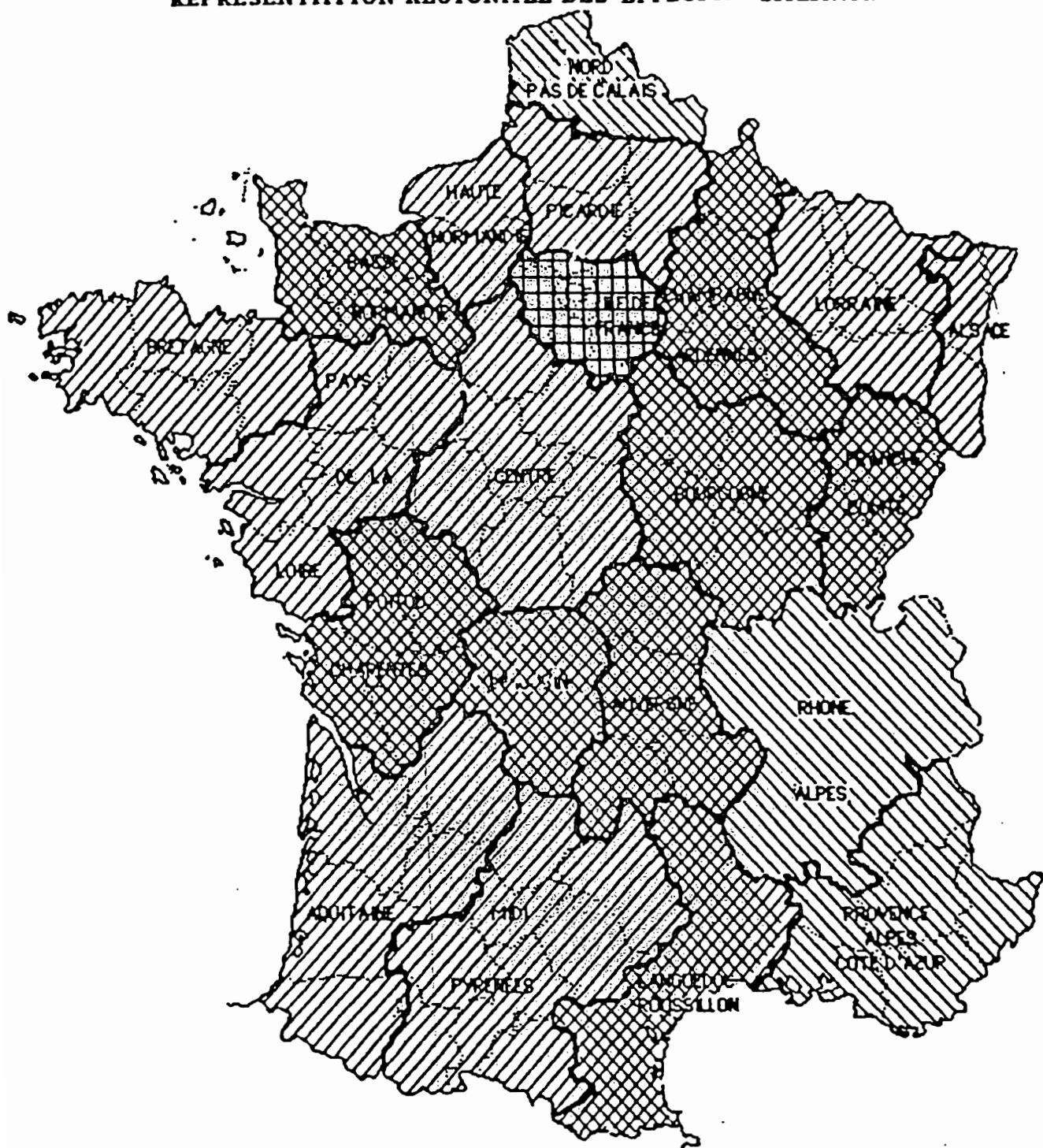
Graphique 9





EVOLUTION DE LA PERIODE AGEE (1975)



Graphique 10

REPRESENTATION REGIONALE DES EFFECTIFS SALARIES



-  régions à très forte représentativité par rapport à la moyenne
-  régions à forte représentativité par rapport à la moyenne
-  régions à représentativité moyenne
-  régions à faible représentativité par rapport à la moyenne



CHAPITRE VII

ANALYSE STRUCTURELLE ET DYNAMIQUE DE L'EMPLOI SALARIE ENTRE 1976 ET 1981

7.0 INTRODUCTION

Le chapitre 6 a permis de faire un tour d'horizon sur quelques résultats statistiques parmi lesquels certains intervenaient à titre illustratif alors que d'autres constituaient le point de départ de l'analyse proprement dite de l'emploi salarié en France au cours de la période 1976-1981.

Il a été constaté, à travers l'analyse des effets simples, que l'emploi salarié par métier, sexe, âge, région est différencié aussi bien en niveau qu'en évolution. Ce résultat, en dépit de son caractère globalisant, n'est pas surprenant dans la mesure où le marché du travail a toujours été cloisonné par les différences dues au sexe, à l'âge, à la profession, à la région, à la formation etc., du moins dans la perspective des théories de la segmentation du marché du travail. Il est d'ailleurs opportun de signaler que c'est dans cette perspective également que se situent les analyses que nous ferons ici.

Toutefois, ce qui reste fondamental, ce n'est pas l'analyse unidimensionnelle de chacune de ces variables, laquelle nous conduirait à une conception néo-classique des mécanismes du fonctionnement du marché du travail, mais plutôt l'analyse de leur interaction puisque ces variables sont étroitement imbriquées et s'influencent mutuellement. On pourra par exemple constater que les différences de l'emploi salarié par métier varient suivant les régions, suivant les classes d'âge, voire le sexe et réciproquement.

Ces différences traduisent généralement l'existence des effets de structure dont l'analyse fera l'objet du présent chapitre. Certes, certains grands résultats observés ailleurs se retrouvent ici, ce qui renforce la cohérence de la méthode que nous avons utilisée. Mais nous situant dans une vision "segmentationniste" du marché, les différents segments de celui-ci ne sont pas établis sur la base des caractéristiques comportementales (comme c'est le cas d'ailleurs des théories duales); ils sont définis à travers les éléments structurels. Cette approche est d'ailleurs en adéquation avec les caractéristiques de l'offre d'emploi, du fait surtout de son caractère structurant.

Ce chapitre développera trois points principaux :

- l'analyse de l'emploi régional à travers les métiers, c'est-à-dire l'analyse de l'effet de structure région-profession. Il se dessinera une typologie des régions, fortement caractérisée

par un certain nombre de métiers qui reflètent plus ou moins la texture économique de chacune d'elle. Ainsi, on retrouve des régions où coexistent de manière importante les métiers du secondaire et du tertiaire (Ile-de-France) ; d'autres à l'inverse sont dominées par les métiers du tertiaire (Provence-Alpes-côte-d'azur). Partant de cette typologie, une analyse détaillée des métiers aux comportements significatifs sera abordée dans quelques régions choisies comme représentatives de chacun des groupes pré-établis. Il apparaîtra que le rôle du tertiaire dans ces comportements est manifeste et c'est pour cela qu'une analyse des facteurs qui gouvernent sa diffusion spatiale sera envisagée.

- le comportement de l'emploi salarié féminin (cf. 7.2) sera analysé au niveau des métiers (cf. 7.2.1) ou au niveau des régions (cf. 7.2.2).

Il a été montré à travers beaucoup d'autres études que le développement de l'emploi féminin apparaît comme «la résultante d'un ensemble d'impulsions provenant de transformations des structures du système capitaliste (structures économiques, sociales, démographiques et institutionnelles) qui ont favorisé l'entrée des femmes dans la vie active» [1]. Notre travail ici se limite aux mutations structurelles d'ordre économique et il nous reviendra de voir dans quelle mesure elles affectent non seulement les métiers, mais également les régions.

L'étude structurelle met en lumière la forte concentration des femmes dans les métiers tertiaires, et qui de surcroît sont d'un très faible niveau de qualification. Bien que ce résultat semble d'une évidence au vu de la discrimination sexuelle du marché du travail, il n'apparaît plus de façon probante à l'échelon régional où l'on ne peut établir un lien de causalité entre féminisation des emplois d'une région et l'importance du tertiaire qu'on y observe. C'est là l'un des résultats intéressants de l'analyse des effets de structure, puisque certains phénomènes, valables au niveau global, s'estompent à une échelle réduite. Cela signifie que d'autres facteurs, outre les mutations des structures économiques, interviennent dans le fonctionnement du marché du travail, et ce au niveau régional, notamment le poids des traditions locales, le dynamisme propre à la région etc.

- la diversité de l'emploi salarié par âge suivant les métiers (cf. 7.3.1) ou suivant les régions (cf. 7.3.2.).

Il existe un certain nombre de métiers dans lesquels excellent les jeunes (les moins 25 ans), catégorie qui mérite une analyse particulière, du fait de la place qu'elle occupe dans le

dispositif d'insertion professionnelle. Il s'agit des métiers à forte rotation de la main-d'oeuvre; ils présentent d'ailleurs un certain parallélisme avec ceux qui ont la faveur des femmes.

Outre cette analyse, il paraît légitime de vérifier si ces résultats sont également valables sur le plan spatial ou interfère une multitude de variables. En les évoquant dans l'analyse, nous terminerons en proposant une typologie régionale de l'emploi salarié par classes d'âge.

7.1 ANALYSE REGIONALE DE LA STRUCTURE DES PROFESSIONS

Le nombre des modalités servant à caractériser les variables profession et région est assez grand (23 modalités pour la variable profession et 21 pour la région). Une analyse plus détaillée de chacune des modalités apparaît plus contraignante dans la mesure où elle nous invite à envisager par exemple l'étude des profils d'évolution des 21x23 effets de paramètres d'ordre 2 de la marge Région x Profession. Certes, une telle étude est riche et permet de rendre plus perceptibles les résultats d'analyse, mais elle mériterait d'être inscrite dans le cadre d'un travail spécifique à développer en profondeur, contrairement à celui que nous menerons ici et qui nous conduira à prospecter les différentes caractéristiques comportementales des salariés à travers les variables que nous avons définies au chapitre précédent.

Nous chercherons à simplifier davantage la nomenclature des professions afin de constituer des groupes de métiers qui ont structurellement une certaine proximité, ce à travers toutes les régions. De manière tout à fait intuitive, on est à même de penser également qu'il existe une certaine typologie régionale suivant les métiers dans la mesure où ceux-ci, dans une région, reflètent au moins partiellement l'activité qui y est dominante, de plus l'activité induit indubitablement des différences spatiales.

Nous serons donc conduits à constituer d'abord des groupes de métiers, lesquels serviront ensuite à caractériser les groupes de régions qui auront été formés. Au-delà, il sera plus justifié de choisir un représentant de chaque groupe de régions sur lequel portera l'analyse à faire, l'accent étant mis sur les métiers qui y sont les plus marqués tant en niveau qu'en évolution. De cette analyse découlera une synthèse d'ensemble quant à la localisation des métiers - plus implicitement des emplois dont ils sont vecteurs - surtout tertiaires pour des raisons que nous évoquerons ultérieurement.

7.1.1 Profils différenciés des professions et des régions

Pour la recherche des groupes de métiers et des groupes de régions , nous utiliserons deux indices respectant la nature multiplicative de la méthodologie exploitée dans le présent travail : l'information discriminante symétrisée [2] et la perte de déviance [3]. Autrement dit, cette recherche est fondée simultanément sur les deux indicateurs et l'idée sous-jacente est la suivante : on regroupe les métiers (respectivement les régions) qui ont des profils voisins toutes les régions confondues (respectivement tous les groupes de métiers confondus) de manière à ce que la perte de déviance (d'information) qui s'en suit soit minimum. L'application de ce principe nous a permis de distinguer cinq principaux groupes de métiers (cf. tableau 22) :

- le groupe I : comprend presque exclusivement des métiers du tertiaire à l'exception des techniciens et dessinateurs qui est un métier qualifié du secondaire. Les deux autres métiers qualifiés du groupe sont les cadres administratifs supérieurs et comptables et techniciens des banques.

- le groupe II : est constitué des métiers du secondaire parmi lesquels, trois sont ceux d'ouvriers qualifiés et un celui d'ouvriers spécialisés et manoeuvres.

- le groupe III s'oppose au groupe précédent puisqu'il est formé uniquement des métiers du tertiaire peu qualifiés à l'exception du personnel qualifié des service para-médicaux et sociaux.

- le groupe IV comprend également des métiers du tertiaire dont trois sont qualifiés (métiers techniques de l'hôtellerie, cadres moyens, Professeurs, professions littéraires et de l'information).

- le groupe V est formé à la fois des métiers du tertiaire (employés non qualifiés, métiers techniques de la distribution) et des métiers du secondaire (ouvriers qualifiés de la radio-électricité et autres spécialités de l'alimentation et autres ouvriers qualifiés).

Tableau 22 : groupe des metiers

GROUPES

Groupe 1	- Mécanographe, opérateur, perforateur
	- Comptables et techniciens des banques
	- Emplés administratifs qualifiés

- Cadres administratifs supérieurs
 - Techniciens et dessinateurs
-

- Groupe II
- Ouvriers qualifiés du travail des étoffes
 - Ouvriers qualifiés des textiles, cuirs et peaux, chimie
 - Ouvriers spécialisés et manoeuvres
 - Ouvriers qualifiés verre, photographie
-

- Groupe III
- Métiers de la santé
 - Personnel qualifié des services paramédicaux et sociaux
 - Métiers des soins personnels
 - Métiers non qualifiés du commerce et distribution
 - Métiers qualifiés de la vente et hôtellerie
-

- Groupe IV
- Employés non qualifiés
 - Ouvriers qualifiés de la radioélectricité, du papier carton et autres spécialités de l'alimentation
 - Métiers techniques de la distribution
 - Autres ouvriers qualifiés
-

- Groupe V
- Cadres moyens
 - Métiers non qualifiés de l'hôtellerie
 - Professeurs, professions littéraires et de l'information
 - Personnel qualifié des techniques comptables
 - Métiers techniques de l'hôtellerie

On a pu également établir cinq groupes de régions (cf. tableau 23) en procédant comme précédemment et chaque groupe a quelques caractéristiques :

- groupe I : formé de l'Ile-de-France uniquement. Ce groupe est caractérisé par une très forte représentativité des métiers de groupe I et une représentativité presque moyenne des métiers du groupe II, la part des techniciens et dessinateurs se situant nettement au-dessus de la moyenne.

- groupe II : comprend des régions de tradition industrielle ancienne telles que le Nord-Pas-de-Calais, la Lorraine, le Rhône-Alpes etc. Les métiers du groupe II sont globalement assez représentés ici et cette représentativité est surtout très forte chez les ouvriers spécialisés et manoeuvres et relativement moyenne chez les ouvriers qualifiés verre,

photographie. Les métiers du groupe I sont à l'inverse très sous-représentés dans ce groupe avec une exception pour le Rhône-Alpes qui a un niveau moyen pour ces métiers.

- groupe III : est constitué de deux régions méditerranéennes que sont le Languedoc-Roussillon et la Provence-Alpes-Côte-d'Azur-Corse. Elles ont une représentativité très largement au-dessus de la moyenne dans les métiers du groupe III et relativement moyenne dans les métiers du groupe II.

-groupe IV : il s'agit des régions de l'ouest où il y a une sous-représentativité des métiers du groupe IV (les métiers techniques de l'hôtellerie restent pourtant très représentés en Bretagne) et une représentativité au-dessus de la moyenne des métiers du groupe III.

-groupe V : comprend des régions telles que le Centre, l'Auvergne, la Franche-Comté etc. où la représentativité des métiers du groupe V est moyenne et celles des métiers du groupe I globalement très au-dessous de moyenne.

Tableau 23 : Groupe de régions

GROUPES	Groupe I	Ile de France
Groupe II	- Champagne-Ardenne	
	- Picardie	
	- Haute Normandie	
	- Bourgogne	
	- Nord-Pas de Calais	
	- Lorraine	
	- Alsace	
	- Rhône-Alpes	
Groupe III	- Languedoc-Roussillon	
	- Provence-Alpes-Côte d'Azur-Corse	
Groupe IV	- Bretagne	
	- Poitou-Charentes	
	- Aquitaine	
	- Midi-Pyrénées	
Groupe V	- Basse Normandie	

- Centre
- Franche Comté
- Pays de loire
- Limousin

Les caractéristiques que nous venons d'attribuer à chacun des groupes de régions restent somme toute assez globalisantes et il serait nécessaire de les affiner.

D'autre part, la typologie obtenue pour les régions reste quasi-stable sur la période 1976 - 1981 bien que des fluctuations moins nettes se manifestent au cours du temps pour certaines régions. On remarquera en passant que les régions les plus représentées dans l'échantillon ont gardé un profil parfaitement stable entre 1976 1981. Peut-on dire, concernant les régions relativement moins représentées qu'il s'agit simplement des fluctuations erratiques à la lumière de simples mouvements qu'elles connaissent au cours de cette période ?

Au vu des caractéristiques des groupes établis, on peut tirer partiellement la conclusion suivante : le degré de représentativité des professions dans une région quelconque est étroitement lié à la structure d'activité dans cette région. Nous essaierons d'apprécier cette conclusion par l'analyse du comportement spécifique de chaque profession au niveau régional, cette appréciation restant limitée à un groupe de régions considérées comme représentatives de chacun des cinq groupes cités ci-haut et sur lesquelles sera focalisée l'analyse. Il s'agit de l'Ile-de-France, Rhône-Alpes, Provence-Alpes-Côte-d'Azur-Corse, Franche-Comté, Bretagne.

7.1.2 Analyse détaillée des professions dans un groupe de régions typées

L'analyse que nous mènerons, même de façon détaillée, ne sera pas centrée sur toutes les professions, l'accent devant être mis sur celles qui ont des caractéristiques particulières aussi bien en niveau qu'en évolution dans une région donnée. Nous analyserons donc les comportements spécifiques des professions en Ile-de-France, en Rhône-Alpes, en Provence-Alpes-Côte-d'Azur-Corse, en Franche-Comté, en Bretagne. Ces régions ont une texture économique soit diversifiée (cas des deux premières avec une diversification plus prononcée en Ile-de-France) soit dominée par le tertiaire (Provence-Alpes-Côte-d'Azur-Corse), soit dominée par le secondaire (Franche-Comté), soit essentiellement agricole (cas de la Bretagne), du moins avant la décentralisation pour cette dernière région.

7.1.2.1 Ile-de-France

L'île-de-France est l'exemple type de la région du groupe I quelle que soit l'année sur la période d'étude que nous avons retenue. Elle reste la seule région dans ce cas. D'ailleurs la "distance structurelle" qui la sépare des autres régions reste relativement grande.

La région est très urbanisée et présente une texture économique assez diversifiée et convenablement équilibrée tant bien dans les activités du tertiaire que celles du secondaire. Serait-ce l'une des explications de l'importante représentativité des salariés dans les métiers du tertiaire et secondaire ?

Toutefois cette quasi-représentativité ne doit pas occulter de nombreuses modifications structurelles - peut-être sont-elles limitées - qui sont apparues avec la crise et également à une période antérieure à la crise avec le processus de décentralisation dont l'un des objectifs était de décongestionner l'hypertrophie de l'île-de-France et qui ont des conséquences sur l'emploi.

Les professions sous-représentées du tertiaire restent ici les métiers de la santé (niveau CAP), les métiers qualifiés de la vente et hôtellerie, les métiers des soins personnels et les métiers non qualifiés du commerce et de la distribution, ces deux derniers métiers se rapprochant tout de même de la représentativité moyenne.

Quant aux professions du secondaire, les métiers d'ouvriers qualifiés du textile, cuirs et peaux, chimie, d'ouvriers spécialisés et manoeuvres et autres ouvriers qualifiés restent sous-représentés. Dans tous les cas, la sous-représentativité concerne ici les professions les moins qualifiées ce qui est presque révélateur du tissu économique de l'île-de-France, qui lui procure en plus une gamme d'emplois qualifiés. A cet égard, on note une très forte représentativité des cadres administratifs supérieurs (ici, ce métier est relativement plus représenté que dans toutes les autres régions) une représentativité également forte des Professeurs, professions littéraires et de l'information (ce sont les métiers les plus qualifiés); d'autres métiers qualifiés tels que métiers techniques de l'hôtellerie, comptables et techniciens des banques, cadres moyens, techniciens et dessinateurs y sont assez représentés.

En évolution, on a enregistré une décroissance des effectifs plus généralement dans les métiers du secondaire à savoir les ouvriers qualifiés verre, photographie, les ouvriers qualifiés de la radioélectricité et les autres spécialités de l'alimentation, les autres ouvriers qualifiés et les ouvriers spécialisés et manoeuvres, ce métier étant le plus sous-représenté en île-de-France

(cf. graphique 11) . Cette baisse reflète la chute des emplois industriels (baisse moyenne de 2,07 % par an de 1975 à 1981) amplifiée par la crise et le mouvement de décentralisation qui s'est fait au détriment de la région.

A l'inverse, les professions qui voient leurs effectifs croître sur la période sont : Mécanographe, opérateur, perforateur malgré un repli entre 1976 - 1977, les métiers non qualifiés de l'hôtellerie, les cadres moyens et dans une moindre mesure les métiers techniques de l'hôtellerie (cf. graphique 12). Plus généralement, ce sont des métiers du tertiaire. Peut-être voit-on là le signe de l'essoufflement du tertiaire en Ile-de-France comme sur le plan national et son évolution la plus favorable limitée aux métiers qualifiés ? Certes, il y a des inflexions au développement du tertiaire en île-de-France (le taux de croissance annuel moyen de l'emploi tertiaire y est passé de 2,53 % sur la période 1968 - 1975 à 1,78 % sur la période 1975 - 1980) mais il assure encore l'essentiel des emplois créés ici. Les différentes évolutions constatées plus haut dans les métiers illustrent de façon grossière le caractère ambivalent de l'offre d'emploi en Ile-de-France avec d'une part l'emploi tertiaire relativement dynamique et d'autre part l'emploi du secondaire qui subit sévèrement les contrecoups de la crise. La résultante de ces deux mouvements se traduit par une chute de l'emploi global dans la région (la diminution du taux de croissance annuel moyen sur la période 1975 - 1980 par rapport à la période 1968 - 1975 est de 1,09 %). Il est étonnant de constater que le rythme annuel moyen de diminution de l'emploi total en Rhône-Alpes est identique à celui de l'île-de-France. Peut-on conclure à une similitude de la structure d'emplois dans les deux régions ou encore que les comportements des métiers y sont identiques ? Seule une analyse des métiers en Rhône-Alpes peut nous apporter des éléments de réponse.

7.1.2.2 Rhône - Alpes

Le Rhône-Alpes est une région de tradition industrielle ancienne de même que les régions du Nord et Nord-Est de France avec qui il forme le groupe II. Cependant, contrairement aux autres régions du groupe, elle a très rapidement développé ses activités du tertiaire puisque le taux de croissance de l'emploi tertiaire retenu comme indicateur de l'importance du tertiaire est de 71 % sur la période allant de 1954 à 1975 [4] , ce taux étant le plus élevé par rapport à celui des autres régions sur la même période. Parallèlement à ce développement du tertiaire, la région a entamé la restructuration de son tissu industriel dans

les années soixante-dix, ce qui fait qu'en proie aux crises de reconversion elle a perdu près de 100.000 emplois industriels entre 1974 et 1984 [5]. Même si par rapport à l'île-de-France, le tertiaire ne connaît qu'un développement récent dans le Rhône-Alpes, les deux régions représentent toutefois des similitudes : leur forte urbanisation d'une part et leur tissu économique où coexistent de façon relativement plus importante le tertiaire et le secondaire. Mais à l'inverse de l'île-de-France, il y a une quasi-représentativité des métiers du secondaire et une quasi-sous-représentativité des métiers du tertiaire en Rhône-Alpes.

Les seuls métiers sous-représentés du secondaire se limitent aux ouvriers qualifiés du travail des étoffes et aux ouvriers qualifiés verre, photographie et ce sont pourtant des métiers qui relèvent des activités dont l'implantation est presque traditionnelle dans la région.

Les ouvriers qualifiés des textiles, cuirs et peaux, chimie, les techniciens et dessinateurs apparaissent comme les métiers les plus représentés et on retrouve encore là les métiers plus implantés dans la région.

Quant aux métiers du tertiaire, bien que sous-représentés, plusieurs se rapprochent de la moyenne, notamment personnel diplômé des services para-médicaux et sociaux, personnel qualifié des techniques comptables, métiers non qualifiés du commerce et distribution, employés non qualifiés, métiers qualifiés de la vente et hôtellerie, comptables et techniciens des banques, cadres administratifs. On peut donc constater plus généralement qu'il n'y a pas de dissémination des métiers comparativement aux autres régions.

En évolution, aucune baisse n'est enregistrée dans les métiers du tertiaire si ce n'est chez les métiers techniques de la distribution où il y a une relative stabilité à partir de 1979. La croissance des effectifs est surtout soutenue et régulière chez les professeurs, professions littéraires et information qui, très relativement sous-représentés en 1976 voient leurs effectifs se situer à un niveau moyen en 1981 (cf. graphique 13). Cette régularité de la croissance des effectifs est également à noter dans les métiers de la santé (niveau CAP), chez les comptables et techniciens des banques malgré un léger repli en 1981 et également chez les mécanographes, opérateurs, perforateurs où l'on remarque une inflexion à partir de 1980.

Quant à la baisse des effectifs, elle se situe surtout chez les ouvriers spécialisés et manoeuvres, et les ouvriers qualifiés de la radioélectricité, du papier carton, qui ont même profil d'évolution : léger accroissement des effectifs entre 1976 et 1977 suivi d'une baisse de

ces effectifs jusqu'en 1981. Ce sont des professions les plus frappées par la crise dans la région.

De l'analyse qui vient d'être faite, l'évolution des métiers est caractérisée par une situation somme toute ambivalente : d'une part la croissance des effectifs dans les métiers du tertiaire et la baisse des effectifs dans les métiers du secondaire. On peut rattacher ces deux mouvements aux caractéristiques de l'appareil de production de Rhône-Alpes que nous avons décrites plus haut et qui se traduisent par une baisse des emplois industriels du fait de la nécessité des restructurations et une hausse des emplois tertiaires dont le taux de croissance, même en-dessus de celui enregistré sur la période 1968 - 1975 (2,61 % de taux de croissance annuel moyen sur la période 1975 - 1980 contre 3,70 % sur la période 1968 - 1975) permet encore de stabiliser, si ce n'est de limiter la contraction de l'emploi global.

Le comportement des métiers dans les deux régions que sont l'île-de-France et le Rhône-Alpes présente grosso modo des liens avec leur appareil de production. En sera-t-il de même pour une région comme la Provence-Alpes-Côte-d'Azur-Corse, dominée par le tertiaire ?

7.1.2.3 Provence Alpes - Côte d' Azur - Corse

Contrairement aux deux régions précédentes, la région Provence Alpes-Côte d'Azur-Corse n'est pas de tradition industrielle ancienne, ce qui lui donne un net avantage dans le développement de ses activités industrielles qui, très récentes, ont été surtout orientées dans les technologies de pointe (informatique et électronique). Cet avantage lui a permis de ressentir de façon moins aigue la crise que les autres régions, l'emploi global augmentant de près de 4,5 % entre 1975 et 1981 (la 3ème région après le Languedoc-Roussillon et le poitou-Charentes), La région a surtout su développer les activités du tertiaire favorisée par sa vocation touristique et sa forte urbanisation.

L'importance du tertiaire et la sous-industrialisation relative de la région nous amènent à tracer un lien avec le comportement des métiers où l'on retrouve une quasi-représentativité des métiers du tertiaire et une quasi-sous-représentativité des métiers du secondaire, ce qui fait de la Provence Alpes-Côte d'Azur-Corse un représentant potentiel des régions du groupe III.

La profession la plus représentée ici est les métiers non qualifiés de l'hôtellerie, ce qui par ailleurs ne semble pas surprenant dans la mesure où la vocation touristique de la région lui

permet de créer assez d'emplois dans ce type de métier. Cette représentativité est la plus forte si on la compare à celle des autres régions dans le même métier.

La représentativité est également assez forte dans les métiers techniques de l'hôtellerie, les métiers non qualifiés de la vente et hôtellerie dont la "proximité" avec les métiers non qualifiés de l'hôtellerie est certaine.

Toutes les professions de la santé et les services marchands retenus dans la nomenclature à 23 métiers sont largement représentés, la situation étant différente de celle que nous avons observée avec l'Ile-de France et le Rhône-Alpes. Certes, il est nécessaire de tempérer l'analyse puisque la région Corse a été agrégée avec la Provence-Alpes-Côte d'Azur, mais cela remet-il profondément en cause les résultats obtenus, du fait du poids relativement faible de la Corse (5,7% dans cette agrégation et la similitude des profils entre les deux régions ?

Les professions du secondaire sont sous-représentées exception faite aux ouvriers qualifiés textiles, cuirs et peaux, chimie, métier relevant des activités dont l'industrialisation récente a permis la modernisation, tout au plus dans les métiers du textile-habillement.

Plusieurs métiers ont des profils d'évolution significatifs :

- les métiers qui marquent une croissance des effectifs sur la période 1976-1977 avant de baisser régulièrement : métiers des soins personnels, métiers techniques de la distribution.
- les métiers dont les effectifs augmentent sur la période 1976-1977 avant de baisser pour s'infléchir à partir de 1980 : personnels qualifiés des techniques comptables, métiers techniques de l'hôtellerie, cadres moyens.
- le troisième profil est celui des métiers où les effectifs croissent jusqu'en 1978 avant de décroître : métiers qualifiés de la vente et hôtellerie, métiers non qualifiés de l'hôtellerie et métiers non qualifiés du commerce et de la distribution (cf. Graphique 16).

Dans les métiers du secondaire, c'est chez les ouvriers qualifiés du travail des étoffes que l'on enregistre une forte progression des effectifs, signe que la crise qui frappe le textile est moins ressentie dans la région du fait de la modernisation de ses industries (cf. Graphique 15).

Ces différents éléments nous amènent à constater que le dynamisme du marché de travail sera surtout stimulé par les emplois administratifs et les emplois liés au tourisme, bref

les emplois tertiaires (l'emploi tertiaire représente 55,8% des emplois en Provence Alpes-Côte d'Azur en 1975) qui néanmoins marquent aujourd'hui un palier (la croissance de l'emploi tertiaire global est de 29% sur la période 1968-1975 contre 15% sur la période 1975-1980).

La Franche comté n'est pas une région aussi tertiarisée que la Provence Alpes-Côte d'Azur et nous allons essayer de voir les caractéristiques comportementales de ses métiers.

7.1.2.4 Franche Comté

La région est plutôt de tradition industrielle ancienne. Elle a comme toutes les autres régions du Nord et Nord-Est enregistré des baisses significatives d'emplois industriels sur la période 1975-1981 (la diminution du taux de croissance annuel moyen est de 1,67%). La région est totalement en situation de mono-industrialisation puisque l'indice de diversification industrielle [6] y est de 24,4 contre 100 pour la France entière.

Le développement du tertiaire y est très moyen depuis les années cinquante. En 1968, le taux d'emploi tertiaire est le plus faible enregistré dans les régions françaises (1358,9 pour 10.000 habitants contre 1793,3 comme moyenne nationale). Sept ans plus tard, même si le tertiaire a augmenté partout, la position de la Franche Comté n'a pas varié puisque le taux d'emploi tertiaire y est toujours le plus bas (1584,3 pour 10,000 habitants ; la moyenne nationale étant de 2043,7). A la même époque, la Franche Comté représente par rapport aux autres régions 1,55% de l'emploi tertiaire l'Ile-de-France en a 26,40% - On aurait pu s'attendre ici, compte tenu de ces caractéristiques à une large représentativité des activités du secondaire. Malheureusement, la région nous révèle une quasi-sous-représentativité de ses métiers du tertiaire et une sous-représentativité relativement plus importante de ses métiers du secondaire, les seuls métiers du secondaire nettement représentés étant les ouvriers spécialisés et manoeuvres, techniciens et dessinateurs et autres métiers qualifiés, ce dernier comprenant comme on le sait un regroupement hétérogène des métiers. Peut-être cette situation s'explique-t-elle par la taille de la région qui est relativement plus faible par rapport aux autres, ce qui fait qu'on peut s'attendre à des fluctuations d'échantillonnage. Les métiers du tertiaire qui y sont les plus sous-représentés sont les comptables et techniciens des banques, et les mécanographes, opérateurs, perforateurs, la sous-représentativité des comptables et techniciens des banques s'opérant de façon spectaculaire sur la période. On remarquera au passage qu'il s'agit d'un métier dont l'activité est liée au développement économique. Donc la sous-représentativité

croissante serait la manifestation de la crise qui a frappé durement aussi bien la région que celles avec qui elle a les mêmes caractéristiques de l'appareil de production.

La croissance des effectifs est ici plutôt favorable aux métiers du secondaire. Tout en étant lente, cette croissance est régulière chez les ouvriers spécialisés et manoeuvres, chez les ouvriers qualifiés verre, photographie où, d'une relative sous-représentativité en 1976 ses effectifs ont largement augmenté sur la période jusqu'en 1981. Les ouvriers qualifiés de la radioélectricité, du papier carton et autres spécialités de l'alimentation et autres ouvriers qualifiés ont un même profil d'évolution : une baisse des effectifs sur la période 1976-1977 (elle est très importante chez les premiers et limitée chez les seconds) suivie d'une hausse régulière des effectifs jusqu'en fin de période.

A l'exception des employés non qualifiés dont les effectifs augmentent régulièrement à partir de 1977, les autres métiers du tertiaire qui ont une évolution significative enregistrent plutôt la baisse de leurs effectifs. C'est le cas des métiers de la santé (niveau CAP), personnel diplômé des services para-médicaux et sociaux, métiers des soins personnels, métiers qualifiés de la vente et hôtellerie, métiers non qualifiés de l'hôtellerie. Ces résultats sont assez parlants en eux-mêmes. Tout d'abord, il s'agit des métiers non qualifiés, ce qui révélerait que lorsque le tertiaire n'est pas dominant dans une région la propension à la baisse est plus sensible dans les métiers les moins qualifiés . D'autre part, tous les métiers de la santé voient leurs effectifs baisser. Est-ce à dire que la politique des pouvoirs publics amorcée depuis 1975 et qui consiste à rationaliser les effectifs de la santé par une politique de restriction budgétaire a trouvé un écho ici plus qu'ailleurs ?

Même avec un tertiaire moins dominant, le marché du travail en Franche Comté n'est pas l'un des plus dégradés et il y est en moyenne plus facile de trouver un emploi [9]. Les taux de chômage y sont faibles ou moyens, mais en forte croissance. Nous terminerons la présentation des régions typées par la Bretagne qui par rapport à la Franche Comté et les autres régions analysées précédemment a une situation économique différente. Elle demeure la région représentative du groupe V.

7.1.2.5 Bretagne

La Bretagne, qui est une région surtout rurale se caractérise par son tissu économique à dominante agricole avec une sous-industrialisation relative dans les années

cinquante. Elle reste néanmoins bénéficiaire de la politique de décentralisation mise en œuvre à cette époque et qui s'est avérée impuissante dans les années soixante-dix, face à la crise de reconversion des régions d'industrie lourde du Nord et Nord-est. L'industrialisation se fera surtout au détriment de Ile-de-France. Le caractère rural de la région n'a pas été un facteur servant de frein au développement du tertiaire. Au contraire, cette activité y est plus présente qu'en Franche Comté où l'urbanisation est relativement moyenne. Le développement du tertiaire est sûrement imputable à la politique de décentralisation. Il reste que la Bretagne est une des régions où l'emploi salarié du tertiaire a le plus augmenté aussi bien sur la période 1968 - 1975 que sur la période 1975 - 1980.

Les professions où l'action des pouvoirs publics est manifeste sont assez représentés ici : métiers de la santé (niveau CAP), avec une très forte représentativité, personnels diplômés des services para-médicaux et sociaux, métiers des soins personnels. Ensuite, viennent les métiers techniques de l'hôtellerie, les métiers qualifiés de la vente et l'hôtellerie, les professeurs, professions littéraires et l'information et les métiers non qualifiés du commerce et distribution, ces deux derniers métiers se situant un peu au-dessus de la moyenne. Ici, le nombre de métiers du tertiaire où il y a une représentativité des effectifs est moyen.

Quant aux professions du secondaire, la représentativité est à noter chez les ouvriers qualifiés de la radio-électricité, du papier carton et autres spécialités de l'alimentation, les autres ouvriers qualifiés et à un degré moindre les ouvriers spécialisés qui ont un niveau moyen de représentativité.

En évolution, la plupart des métiers n'ont pas d'évolution significative. Cependant, les effectifs augmentent chez les employés non qualifiés, les cadres administratifs supérieurs, et les métiers de la santé (niveau CAP).

La baisse des effectifs est très spectaculaire chez les mécanographes, opérateurs, perforateurs où l'on passe d'une représentativité moyenne en 1976 à une forte sous-représentativité en 1981. Rien ne semble justifier ce phénomène.

On relève également une baisse des effectifs chez les cadres moyens et les autres ouvriers qualifiés où les effectifs décroissent d'une manière régulière à partir de 1977 avant de

se stabiliser entre 1980 et 1981 ; cette baisse est à signaler également chez les métiers qualifiés vente et hôtellerie.

Cette étude, limitée aux cinq régions seulement, permet d'éclairer au mieux la diversité et la complexité spatiales des comportements des métiers. Si à la suite des grandes concentrations industrielles, on a pu vérifier une certaine "polarisation des emplois", à l'heure du tertiaire, on assiste plutôt à la propension à une dissémination spatiale des emplois de ce secteur, ce qui reste également une preuve de sa diversité. Celle-ci est d'autant plus grande que deux régions aux caractéristiques voisines et au profil de représentativité proche présentent plutôt des évolutions contrastées. C'est le cas par exemple des métiers techniques de l'hôtellerie, aussi bien représentés en île-de-France qu'en Provence-Alpes-Côte d'Azur-Corse, deux régions au niveau d'urbanisation élevé et où les évolutions dans ces métiers sont plutôt antinomiques : croissance de ses effectifs en île-de-France et leur décroissance en Provence-Alpes-Côte d'Azur-Corse. Il ressort alors que, outre l'activité, le dynamisme d'une région jouerait un rôle non moindre dans la structuration des métiers. Mais comment appréhender le dynamisme régional dans la diversité des métiers ? La réponse à cette question n'est pas évidente et nécessite au moins une étude individuelle de chaque métier dans chaque région, ce qui serait encore plus contraignant.

Dans les régions que nous avons choisies ici, les évolutions les plus significatives concernent les métiers du tertiaire. Outre le fait que la nomenclature qui a été constituée comporte une forte proportion des métiers du tertiaire, ces évolutions traduisent sûrement le rôle qu'il joue désormais dans la dynamique de l'emploi national comme régional. Peut-on voir dans les clivages qui apparaissent dans leurs évolutions les facteurs déterminant la localisation du tertiaire dans une région ?

7.1.3 Déterminants de la localisation du tertiaire

La prospection des facteurs déterminant le tertiaire dans une région nous conduit à en distinguer deux types:

- les facteurs extra-économiques,
- les facteurs conjonctuels.

L'indicateur de la localisation du tertiaire dans une région reste l'emploi tertiaire bien évidemment.

7.1.3.1 Facteurs extra-économiques

L'évolution de l'emploi tertiaire régional [8] montre que, malgré son ralentissement sur la période 1975 - 1981, l'emploi tertiaire a continué ses tendances antérieures - elles sont caractérisées par une croissance continue des effectifs - avec des inflexions limitées. On peut lier ces évolutions à des facteurs tels que l'urbanisation, l'action des pouvoirs publics et à un degré moindre la structure des qualifications des emplois.

Le degré d'urbanisation détermine au moins partiellement l'emploi tertiaire [9] : « avec celle-ci (urbanisation), la demande des services augmente et se diversifie ; par ailleurs, des déséconomies externes peuvent apparaître avec l'augmentation de la taille des agglomérations ; ceci nécessite des emplois nouveaux, en particulier pour l'administration locale ou des activités comme les services aux entreprises et aux particuliers ». C'est donc ce facteur qui expliquerait le niveau de l'emploi tertiaire en île-de-France, Provence-Alpes-Côte d'Azur-Corse, Rhône-Alpes (en 1975, le tertiaire représentait 61,7% de l'emploi total en île-de-France et 55,8 % en Provence-Alpes-Côte d'Azur-Corse). Cependant, le facteur urbain, même s'il induit l'emploi tertiaire voire sa localisation n'apparaît plus de façon probante, le cas de la Bretagne étant à cet égard symptomatique. La région qui est plutôt à dominante rurale a un emploi tertiaire plus soutenu qu'en Franche-Comté dont le niveau d'urbanisation, bien que relativement moyen reste supérieur à celui de la Bretagne. En somme, l'emploi tertiaire qui dans les années soixante-dix se développait prioritairement dans les grandes métropoles s'étend sur la période 1975 - 1981 aux régions rurales, peu urbanisées voire peu développées. C'est le cas des régions telles l'Auvergne, le Limousin, le Poitou-Charente, la Basse-Normandie les deux dernières sont celles de l'ouest qui ont bénéficié de la décentralisation bien qu'il s'agisse sûrement d'un effet de "rattrapage" dans la mesure où ces régions qui ont occupé les dernières positions en matière d'emploi tertiaire sur la période 1954 - 1975, se situent à une position très favorable entre 1975 - 1981, période au cours de laquelle le tertiaire s'essouffle dans presque toutes les autres régions.

La politique des pouvoirs publics a également une influence sur la localisation du tertiaire : « par ses choix en matière de sécurité sociale, l'Etat joue un rôle dans les métiers de la santé » [10]. L'emploi induit par les pouvoirs publics à une diffusion beaucoup plus homogène dans l'espace. C'est le cas par exemple des métiers de la santé où les différences

régionales ne sont pas très importantes par rapport aux autres métiers. Cette diffusion spatiale relativement homogène des métiers de la santé s'oppose à la polarisation des métiers les plus qualifiés qu'on retrouve en île-de-France, Provence-Alpes-Côte d'Azur - et à un degré moindre Rhône-Alpes : ce sont les cadres administratifs supérieurs, les cadres moyens, les professeurs, professions littéraires et de l'information etc. qui constituent ce que l'on appelle le tertiaire supérieur. Donc, la qualification détermine le tertiaire ou du moins la polarisation de l'emploi tertiaire. Ces éléments extra-économiques devraient être complétés par d'autres facteurs beaucoup plus conjoncturels dont on a déjà vu quelques traits ici et qui ont également une influence sur la localisation du tertiaire.

7.1.3.2 Facteurs conjoncturels

On peut dire de façon plus générale que les évolutions de l'emploi tertiaire sont en partie déterminées par celles de l'industrie. En effet, l'industrie dont la répartition est très hétérogène sur le territoire a une influence sur le tertiaire. Les régions du Nord et Nord-Est ont subi de nombreuses pertes d'emplois industriels sur la période récente et c'est dans ces régions que l'évolution de l'emploi tertiaire reste très faible. D'autre part, l'emploi tertiaire augmente fortement dans les régions où l'emploi industriel augmente également. C'est le cas des régions de l'Ouest telles que la Bretagne, le Poitou-Charente, le Pays de la Loire où se font sentir les effets de la décentralisation. Il y a donc un lien qui apparaît entre l'emploi industriel et l'emploi tertiaire et celui-ci reflète à peu près les évolutions conjoncturelles, diversement ressenties sur le plan géographique. C'est la demande qui gouverne ces évolutions et on peut à cet égard distinguer comme Lipietz [11] deux tertiaires :

- le tertiaire "archaïque" qui est la petite unité de production artisanale ou le petit commerce et où la relation de proximité nécessaire avec la demande est très forte La diffusion spatiale de ce tertiaire est assez grande et son implantation dans une région serait le reflet plus ou moins déformé de la répartition spatiale de la population et on pourrait parler alors de "tertiaire démographique". Cette implantation est d'ailleurs très importante dans les régions dotées d'une très grande métropole où la demande est forte vu l'importance de la population. L'essoufflement du tertiaire dans l'ensemble des régions peut au moins être imputé à la baisse de la demande induite par la crise.

- le tertiaire "moderne" qui remplace de plus en plus le tertiaire "archaïque". Il y a ici une introduction des nouvelles techniques dont la conséquence est la modification du processus de travail. Ce tertiaire est relativement plus "libre" par rapport à la demande et l'utilisation des techniques nouvelles peut provoquer la concentration des emplois dans certains endroits du territoire, en l'occurrence les emplois qui exigent une main-d'oeuvre qualifiée.

Ces éléments nouveaux nous amènent à relativiser l'autonomie du tertiaire par rapport à la crise. Ce secteur ayant assuré l'essentiel des emplois depuis 1974, il est fort à craindre que la prorogation de la crise ne vienne mettre en cause son rôle dans l'atténuation des effets de cette crise, accentuant ainsi les tendances de son essoufflement déjà constatées sur la période 1975 - 1981, ce qui serait encore lourd de conséquences pour des régions comme le Nord-Pas-de-Calais, la Lorraine etc. où la manifestation de la crise est dominante.

L'extrême diversité des métiers au niveau régional a été mise en relief par l'analyse que nous venons de faire. La prise en compte de la variable sexe aurait permis de mieux éclairer cette diversité, la profession étant très fortement liée au sexe comme le montrent les statistiques de choix des modèles (cf. 6.2.1).

D'autre part, le développement de l'emploi féminin est un phénomène qui semble insensible à la crise et au-delà des transformations qui le modulent, il devient nécessaire d'explorer ses caractéristiques, c'est-à-dire les métiers qui ont la faveur des femmes. Une fois effectuée cette exploration, on pourra également analyser le comportement des femmes à travers les régions afin de juger si les différences régionales restent tributaires des comportements spécifiques ou des métiers eux-mêmes.

7.2. PARTICIPATION FEMININE PROFESSIONNELLE ET REGIONALE

La féminisation de l'emploi, amorcée dans les années soixante, n'a pas connu de repli jusqu'aux années quatre-vingts, nonobstant la rupture intervenue en 1974 dans la croissance économique. Au contraire, elle s'est même amplifiée dans la crise, son ampleur étant surtout manifeste chez les femmes de 25 - 34 ans qui enregistrent un taux de participation à l'activité de 63,5 % en 1982 contre 53,1 % en 1975 et 44,6 % en 1968 [12]. Ainsi, même avec une conjoncture dégradée les femmes sont de plus en plus présentes sur le marché du travail.

Certes, on peut noter sur la période 1975-1980 une légère diminution des taux d'activité des femmes de moins de 23 ans et celles de plus de 55 ans imputable à l'allongement de la scolarité pour les premières et l'abaissement de l'âge de la retraite pour les secondes, mais cela ne met aucunement en cause les tendances générales.

La croissance de l'activité féminine est à lier partiellement à l'extension de la salarisation, qui fait appel à une main-d'oeuvre de réserve dont les femmes ne sont qu'une des composantes. Cette salarisation est en aval de différentes transformations du processus voire de l'appareil de production en cours durant les deux dernières décennies et qui requièrent très généralement une main-d'oeuvre aux valeurs "féminines". Bref, une multitude de facteurs possibles déterminent l'évolution récente de l'activité féminine, notamment les facteurs culturels, sociologiques, institutionnels, économiques, géographiques et le statut professionnel etc. Nous n'aborderons pas tous ces facteurs ici, nous limitant tout simplement à analyser l'influence des professions et des caractéristiques régionales sur la participation féminine entre 1976 et 1981. Une telle analyse sera menée en deux temps avant de rechercher le rôle que jouent les métiers dans la structuration de la participation féminine régionale. Ainsi, sommes-nous conduits à analyser successivement :

- la participation féminine professionnelle,
- la participation féminine régionale,
- le rôle déterminant des métiers dans la participation féminine régionale,

7.2.1 Participation féminine professionnelle

Le métier est un élément de différenciation assez net de la participation féminine. Ainsi, on note une importante présence féminine dans les métiers de personnel de secrétariat (98 %), d'ouvriers qualifiés du travail des étoffes (86 %) contrairement aux métiers d'ouvriers qualifiés de la mécanique ou les autres ouvriers qualifiés, métiers dans lesquels cette participation est inférieure à 1 % [13]. Nous apprécierons la participation des femmes à l'emploi à travers la nomenclature à 23 métiers constituée ici. Une fois esquissée la typologie d'ensemble de cette participation, l'accent sera mis sur son évolution au cours de la période 1976 - 1981.

7.2.1.1 Typologie d'ensemble : une polarisation de la participation féminine

Le lien entre le développement du tertiaire et la croissance de l'activité féminine a été déjà implicitement évoqué précédemment. Ce lien est d'autant plus fort que le secteur tertiaire concentre une part importante de femmes salariées (75 % de femmes salariées travaillent dans le tertiaire en 1981). Le recours à la nomenclature à 23 métiers permet de constater que près d'un métier sur deux à une très forte participation féminine. De plus, sur dix métiers à très forte participation féminine, neuf appartiennent au tertiaire. On peut sur la base de cette nomenclature dégager trois groupes de métiers suivant que les femmes y sont fortement ou peu présentes :

- les métiers où la participation féminine relative par rapport à la moyenne est très forte. C'est le cas des métiers tels que : métiers de la santé, ouvriers qualifiés du travail des étoffes, personnel diplômé des soins para-médicaux et sociaux, métiers des soins personnels, personnel qualifié des techniques comptables, mécanographe, opérateur, perforateur, métiers non qualifiés du commerce et de la distribution, employés non qualifiés, métiers qualifiés de la vente et hôtellerie, métiers techniques de l'hôtellerie.

- métiers où la participation féminine relative est moyenne : ce sont les métiers non qualifiés de l'hôtellerie, les employés administratifs qualifiés, les ouvriers qualifiés cuirs et peaux, chimie, et à un degré moindre les ouvriers spécialisés et manoeuvres et un ensemble de métiers qualifiés voire hautement qualifiés tels que cadres moyens, comptables et techniciens des banques, métiers techniques de la distribution, professeurs, professions littéraires et de l'information.

- métiers où la participation féminine relative est faible par rapport à la moyenne : on retrouve ici les métiers tels que cadres administratifs supérieurs, techniciens et dessinateurs qui sont assez qualifiés et d'autres métiers moins qualifiés notamment ouvriers qualifiés de la radio-électricité, du papier carton et autres spécialités de l'alimentation et autres ouvriers qualifiés, ces derniers étant tous des métiers du secondaire.

À la lumière de cette typologie se dégagent quelques idées fortes. En effet, il y apparaît une structure bipolaire des métiers, marquée par une forte présence des femmes dans les métiers du tertiaire et une forte présence d'hommes dans les métiers du secondaire. De plus, si l'on s'en tient uniquement au pôle où elles sont fortement représentées, on constate que les femmes occupent des emplois peu qualifiés. D'ailleurs, la présence des femmes dans un

métier est inversement liée au degré de qualification de ce métier. Ainsi, tous les métiers du groupe des métiers où la participation féminine est relativement forte par rapport à la moyenne sont peu qualifiés à l'exception de quelques métiers qualifiés tels que personnel qualifié des soins para-médicaux et sociaux et les métiers techniques de l'hôtellerie. Le groupe des métiers à participation féminine relativement moyenne comprend quant à lui soit des métiers du tertiaire qualifiés (cadres moyens, comptables et techniciens de banques etc.) soit d'autres métiers peu qualifiés du tertiaire et du secondaire. On peut donc dire que la participation féminine est liée non seulement au métier, mais également au niveau de qualification de ce métier.

Au-delà de ces premiers éléments d'appréciation, il serait intéressant d'analyser la dynamique de la féminisation des professions afin de vérifier d'une part si la structure esquissée ci-haut reste valable dans le temps, d'identifier d'autre part les métiers à évolution significative.

7.2.1.2 Evolution de la participation féminine professionnelle sur la période 1976 - 1981

Sur la période 1976 - 1981, ce sont en général, les métiers du tertiaire qui enregistrent des évolutions significatives sur le plan de la féminisation des emplois. L'examen de la discrimination sexuelle relative montre que seuls deux métiers sur les douze qui ont une évolution significative sont du secondaire. Nous distinguerons les métiers où cette discrimination sexuelle relative évolue favorablement (cf. graphique 19) de ceux où son évolution est défavorable (cf. graphique 20) sur la période 1976 -1981.

L'examen du graphique 19 indique une évolution spectaculaire de la féminisation relative des métiers de la santé au cours de la période d'étude. Cette féminisation a été régulière et reste imputable à la quasi-stabilité de la féminisation absolue conjuguée à une baisse régulière et forte des effectifs masculins dans ces métiers (cf. graphique 21). Il s'agit donc d'un métier très féminisé (huit femmes sur dix environ y exercent une activité) et entre 1974 et 1980, malgré la croissance des effectifs dans ce métier (7% de taux de croissance annuel moyen), il est resté de plus en plus fermé aux hommes. D'ailleurs, 320.000 emplois féminins y ont été créés.

Les employés non qualifiés, déjà très féminisés également connaissent comme les métiers de la santé une féminisation relative en progression et elle est liée à la masculinisation absolue qui après une période stable jusqu'en 1978 a commencé à baisser. De même la féminisation absolue de ce métier est restée relativement stable. Il faut noter que la progression de cette féminisation relative est de moindre importance comparée à celle des métiers de la santé.

Les employés administratifs qualifiés sont dans une situation analogue à celle des métiers précédents, mais cette fois-ci, on assiste à une baisse des effectifs masculins conjuguée à une féminisation absolue croissante et l'inflexion entre 1977 et 1979 traduit la légère baisse des effectifs féminins dans cet intervalle de temps.

D'autres métiers tels que les cadres administratifs supérieurs, les cadres moyens et les métiers techniques de l'hôtellerie voient une progression de leur féminisation relative aussi sur la période 1976-1981 et cette évolution est surtout le fait de l'évolution favorable de la féminisation absolue.

Le premier constat qui découle de cette analyse est que la féminisation relative touche surtout les métiers du tertiaire soit très féminisés et peu ou moyennement qualifiés. soit peu féminisés mais à haut niveau de qualification. Dans ce dernier cas les pouvoirs publics jouent un très grand rôle notamment en matière de création d'emplois féminins qualifiés dans l'administration et dans l'enseignement.

A l'inverse du mouvement que nous venons de décrire, d'autres métiers connaissent plutôt une déféminisation relative (cf. Graphique 20).

L'évolution des ouvriers qualifiés de la radioélectricité, du papier-carton et autres spécialités de l'alimentation est très significative et reflète un peu le caractère de ce métier fermé aux femmes. La déféminisation absolue est très forte (cf. Graphique 21) et les effectifs masculins bien que stables en début de la période ont tendance à augmenter en fin de période. Tout laisse à croire qu'avec la crise qui frappe le secondaire en général, les femmes sont les plus touchées dans ce métier.

Certains métiers du tertiaire tels que les métiers qualifiés de la vente et de l'hôtellerie ont une déféminisation relative imputable également à une déféminisation absolue, conjuguée à une masculinisation absolue qui reste néanmoins limitée.

Peuvent également faire partie de ces métiers à déféminisation relative, les métiers tels que le personnel qualifié des techniques comptables, les ouvriers spécialisés et manoeuvres où cette déféminisation s'amorce en 1977 même si elle reste lente et dans une moindre mesure les mécanographes, opérateurs, perforateurs où il y a globalement une baisse des effectifs féminins entre 1976 et 1981.

On constate à la suite de cette analyse que les différentes évolutions ne mettent pas tellement en cause la typologie esquissée en (2.2.1.1), celle-ci restant quasiment stable sur la période 1976 - 1981. Autrement dit, les femmes ont continué à investir les métiers du tertiaire et il existe un cloisonnement toutefois, induit par le degré de qualification et se traduisant par une polarisation des femmes dans les métiers les moins qualifiés de ce secteur. Il aurait bien fallu rechercher les raisons qui gouvernent cette polarisation et font que certains métiers restent fermés aux femmes.

Tous ces métiers où les femmes sont largement représentées sont jugés presque féminins à cause de leurs caractéristiques : présentation , accueil dans les métiers tels que métiers non qualifiés du commerce et distribution, employés non qualifiés de la vente et hôtellerie etc., précision dans les métiers d'ouvriers qualifiés du travail des étoffes, affection et attention dans les métiers de la santé. Ces caractéristiques sont les qualités que possèdent généralement les femmes et elles sont souvent acquises dans la pratique des tâches ménagères ou sont inhérentes à la nature féminine [14]. Même lorsque les femmes investissent des métiers qualifiés, ce sont ceux où les valeurs traditionnelles féminines restent dominantes, l'exemple du personnel diplômé des soins para-médicaux et sociaux étant révélateur. On peut alors dire que l'accès des femmes à des emplois qualifiés n'est que "la projection dans la sphère marchande du travail domestique traditionnellement réservé aux femmes".[15]

La formation des femmes explique également leur présence forte dans certains métiers précis et les jeunes filles sont même «confrontées à des problèmes d'insertion professionnelle lorsqu'elles ont été formées à des emplois non traditionnellement féminins» [16] ; sur le plan promotionnel, on sait que les emplois d'encadrement plus qualifiés en général se font très souvent par le truchement des stages de formation professionnelle, lesquels sont peu ouverts aux femmes pour des raisons familiales d'une part ou qui tiennent des politiques de gestion de main-d'oeuvre des entreprises d'autre part. On peut invoquer d'autres raisons telles que les conditions de travail, l'organisation du travail qui, lorsqu'elle est parcellisée - travail taylorien - offre des emplois peu qualifiés aux femmes ; ces emplois sont qualifiés lorsqu'il y a

automatisation du processus de production qui implique en plus le travail de nuit, un autre élément de frein au recours à la main-d'oeuvre féminine. Ces raisons sont aussi géographiquement vérifiées et on peut donc s'arrêter un instant pour étudier le comportement en matière d'emploi des femmes au niveau régional.

7.2.2 Participation féminine régionale

L'examen des effets de 1er ordre (cf. 6.3) nous a révélé que la région avait un pouvoir de discrimination moins important que la profession. Ce résultat reste valable lorsque nous associons la variable sexe à chacune des deux variables. En d'autres termes, l'hétérogénéité des comportements régionaux en matière d'emploi féminin est moins prononcée que celle induite par les métiers.

L'un des résultats observés dans le précédent paragraphe (cf. 7.2.1) est la forte concentration des femmes dans les métiers du tertiaire. Cette prédominance des emplois tertiaires dans les emplois occupés par les femmes au niveau national se retrouve au niveau régional et quelle que soit la région, plus de 50 % des femmes travaillent dans le tertiaire (52 % en Franche Comté, région la moins tertiariée contre 75 % en Provence-Alpes-Côte d'Azur, région la plus tertiariée). Ce phénomène peut être lié au développement du tertiaire qui s'est étendu à toutes les régions dans la décennie soixante-dix, et apparaissant comme un facteur d'atténuation des disparités régionales, même si ce mouvement a été en partie limité par la crise. D'autre part, il serait fâcheux, au-delà de ces résultats, de présumer de l'existence d'une relation causale entre tertiarisation d'une région et forte participation féminine à l'activité. Ce serait d'ailleurs occulter les éléments structurels, voire spécifiques qui modulent les comportements des femmes à l'activité dans une région.

L'objet de ce sous-paragraphe, comme le précédent, sera de prospecter les différents pôles spatiaux en matière de participation féminine d'une part, et les régions où le mouvement des femmes sur le marché du travail est marqué sur la période d'étude d'autre part. Nous terminerons en précisant le rôle des métiers dans la structuration de l'emploi féminin régional.

7.2.2.1 Participation féminine : des comportements régionaux relativement diversifiés

Dans une région donnée, la participation féminine à l'activité dépend de plusieurs facteurs qui interfèrent les uns sur les autres : la structure démographique, la qualification, le taux d'activité, la structure des emplois etc. Cependant, la réalité est beaucoup plus complexe, et compte tenu de l'analyse effectuée au paragraphe 7.1, il apparaît que les femmes ne sont pas toujours assez représentées où la structure des emplois leur semble favorable. D'autre part, deux régions ayant une "proximité structurelle" peuvent avoir des comportements en matière de participation féminine totalement différents. C'est qu'il existe des caractéristiques spécifiques à chaque région, lesquelles contribuent à moduler également leur processus de féminisation. Ainsi que nous l'avons fait pour les professions, on peut distinguer trois grands types de régions du point de vue de la féminisation des emplois sur la période 1976 - 1981, même si en évolution certaines régions comme l'Aquitaine mettent partiellement en cause cette typologie :

- régions où la féminisation relative se situe au-dessus de la moyenne nationale. C'est le cas notamment des régions telles que la Franche Comté, l'Ile-de-France, la Bourgogne. Peuvent également faire partie de ce groupe la Haute-Normandie et les Pays de la Loire. Pour ce qui est de l'Ile-de-France, on peut expliquer cette forte représentativité des femmes grâce aux emplois tertiaires offerts qui y jouent un rôle attractif sur la main-d'oeuvre féminine. De plus, la scolarité est assez longue dans la région ; la conjonction d'un haut niveau de formation et d'une sous-fécondité dans la région favorise un fort taux de maintien à l'activité. En Franche Comté, la forte féminisation des emplois est surtout liée au fort taux de féminité dans l'industrie où travaillent près de 40 % des femmes contre une moyenne nationale qui n'est que de 26 % : ce taux avoisine 30 % dans les régions du Bassin Parisien.

- régions où la participation féminine relative est à niveau moyen (moyenne nationale) : Rhône-Alpes, Basse-Normandie, Auvergne, Picardie, Champagne-Ardenne, Poitou-Charentes, Bretagne. On constate qu'il s'agit ici des régions, soit très anciennement industrialisées, soit agricoles. Dans ce dernier cas, on sait que la féminisation des emplois s'est surtout opérée dans le secondaire du fait de la décentralisation industrielle qui a créé beaucoup d'emplois, mais dans le cadre d'une banalisation des travaux, laquelle a fourni d'importantes opportunités d'emploi aux femmes.

- régions où la participation féminine relative est en-deçà de la moyenne. Ce groupe comprend : Alsace, Nord-Pas-de-Calais, Lorraine, Limousin, Mid-Pyrénées, Languedoc-Roussillon, Provence-Alpes-Côte d'Azur, Aquitaine. On peut schématiser en disant que ce groupe comprend les vieilles régions industrielles du Nord-Est et les régions tertiaires du sud. Les premières sont caractérisées par l'existence d'une proportion importante d'emplois industriels éprouvants (le textile utilise une main-d'oeuvre jeune, capable de rythmes de travail élevés et d'une grande dextérité), ce qui est un frein à la féminisation des emplois. D'autre part, il s'agit des régions où la fécondité est assez élevée. La plupart des femmes commencent à travailler très jeunes et s'arrêtent à l'occasion d'un mariage ou des premières naissances.

A l'inverse, la situation n'est pas identique dans le sud où le tertiaire prédominant devrait favoriser la féminisation des emplois. On peut imputer cela en partie à la situation défavorable du marché du travail, puisque depuis le début des années soixante-dix, ces régions enregistrent de très forts taux de chômage.

Il serait hasardeux d'associer la sous-activité féminine des régions du Sud aux facteurs démographiques. En effet, la fécondité y est des plus faibles et cela ne contribue pas à favoriser la féminisation des emplois. Au-delà, on peut évoquer d'autres facteurs tels que les traditions régionales. En ce sens, les régions du Sud restent encore attachées au modèle traditionnel de "la femme à la maison".

Cette typologie éclaire mieux les effets spécifiques régionaux sur la participation féminine. Les indicateurs tels que la fécondité ou la tertiarisation des emplois ne suffisent pas pour caractériser la féminisation des emplois dans une région. A cet égard, un paradoxe apparaît ici et qui est révélateur des spécificités régionales : la Provence-Alpes-Côte d'Azur a une structure d'emploi se caractérisant par la quasi représentativité des métiers du tertiaire (cf. 7.1.2.3), ce qui suppose que l'on s'attende à y enregistrer une très forte participation féminine, les femmes étant sur le plan global (sur le plan national) largement concentrées dans le tertiaire (cf. 7.2.1). Malheureusement, il n'en est pas le cas et la région se singularise par sa sous-féminisation relative. Bien que huit femmes sur dix exercent une activité dans le tertiaire, dans cette région, le taux de féminisation des emplois y reste l'un des plus bas à l'échelle nationale. Malgré son caractère attractif, du fait de sa structure des emplois, d'une part et l'évolution favorable des créations d'emploi sur la période 1975 - 1981 (4,5 % d'augmentation global sur cette période) la région n'a pas pu drainer vers elle des actifs

féminins, ce qui semble normal dans la mesure où les migrations sont liées à l'âge et au sexe, les femmes quittant de moins en moins leur région d'origine [17].

En revanche, la Franche Comté, région la moins tertiaisée se caractérise par la participation féminine relative la plus forte. La quasi-totalité des métiers du tertiaire y sont sous-représentés (cf. 7.1.2.4). De plus, la féminisation du tertiaire y est la plus faible en 1981 (63 % contre 75 % sur le plan national et 86 % en Provence-Alpes-Côte d'Azur).

Ces observations nous incitent à réfuter toute relation positive entre féminisation d'une région et tertiaisation de celle-ci, ainsi qu'on a pu l'observer sur le plan national. Le lien semble être celui d'une association de la féminisation d'une région à son industrialisation. Nous y reviendrons.

La question qu'on peut se poser est de savoir si l'évolution de la participation féminine relative sur la période de crise a été limitée ou pas. On peut globalement constater que les régions qui en 1976 étaient relativement plus féminisées le sont davantage en 1981, ou encore sont restées relativement stables. Celles qui étaient peu féminisées le sont encore moins.

7.2.2.2 Evolution régionale de la participation féminine entre 1976 et 1981

Sur la période 1976 - 1981, les régions à forte participation féminine relative ont une évolution favorable du point de vue de la féminisation des emplois. En effet, c'est au Centre et en Franche Comté que l'on observe une féminisation des emplois de plus en plus nette sur la période, malgré une inflexion en fin de période (cf. graphique 23).

Le Centre, au cours de la période a vu ses effectifs salariés globaux augmenter. Cette augmentation a été surtout bénéfique aux femmes dont la croissance régulière des effectifs (avec une inflexion en 1981) conjuguée à une relative stabilité des effectifs masculins sur la même période (cf. graphique 22a et 22b) ont renforcé l'évolution assez positive de la féminisation relative.

En Franche Comté, l'évolution de la féminisation relative est à imputer à un mouvement qui est le résultat de deux tendances antagonistes : l'une caractérisée par la progression des effectifs féminins jusqu'en 1979 avant de s'estomper, voire se stabiliser et l'autre caractérisée quant à elle par une baisse régulière des effectifs salariés sur la période 1976 - 1979 avant d'augmenter légèrement jusqu'en fin de période.

L'Ile-de-France et la Bourgogne qui se situent parmi les régions les plus féminisées ont un même profil d'évolution et qui se caractérise par une stabilisation de la féminisation relative sur la période 1976-1981. Cette situation s'explique par la stabilité des effectifs féminins et masculins au cours de cet intervalle de temps, même si l'on note une atténuation des effectifs féminins dans les deux régions entre 1979 et 1980 et que reflète la courbe de la féminisation relative.

L'évolution de la féminisation relative en Basse Normandie présente des similitude avec celle des deux régions que sont L'Ile-de-France et la Bourgogne.

Il faut noter que l'évolution la plus forte de la féminisation relative est à mettre à l'actif de la région Aquitaine, déjà assez masculinisée. D'un niveau relativement faible en début de période, la féminisation relative est passée à un niveau moyen en fin de période et ce mouvement est le résultat de deux tendances antinomiques : une croissance régulière des effectifs salariés féminins entre 1977 et 1981 alors que dans le même temps, la part des hommes dans l'effectif total baisse. La féminisation relative serait à lier ici à la féminisation du tertiaire (5% en évolution sur la période 1975-1981) et il n'y a pas de faiblesse généralisée dans les différents métiers du tertiaire en Aquitaine.

En revanche, l'évolution défavorable de la féminisation relative en Lorraine s'expliquerait par la faible tertiarisation (certes on a montré plus haut qu'il n'y avait pas de lien qui apparaisse de manière tranchée entre le niveau du tertiaire et le niveau de l'emploi féminin dans une région) conjuguée à une spécialisation industrielle qui reste un frein au développement de l'emploi féminin. Même dans certains métiers comme le textile-habillement qui requiert généralement une main-d'oeuvre féminine ou encore l'hôtellerie et le commerce, l'emploi est en déclin et les femmes, sont les premières victimes de cette situation. Il y a une évolution en deux temps des effectifs ici : progression de l'emploi féminin entre 1976 et 1979 (la croissance est même très forte entre 1977 et 1978) suivie d'une baisse très forte jusqu'en fin de période. L'emploi masculin a une évolution symétrique à celui des femmes, mais de moindre importance.

Bien que de tradition industrielle, on aurait pu s'attendre à une situation plus que favorable de la participation féminine en Alsace et ce du fait de son urbanisation relative de même que la construction européenne qui a profité au tertiaire. La région enregistre plutôt une déféminisation relative et, compte tenu de la structure de l'emploi qui semble favorable aux

femmes, on peut penser que la structure démographique a modulé cette évolution. Toutefois, l'emploi féminin regresse ici alors que l'emploi masculin reste relativement stable sur la période 1976 - 1981.

Ces évolutions contrastées, au-delà des spécificités régionales qui les déterminent en partie, semblent plus tributaires d'un élément fondamental qui joue un grand rôle dans le processus de féminisation des emplois au niveau d'une région : les professions. Dans quelle mesure celles-ci peuvent-elles alors permettre d'élucider la diversité des participations féminines régionales ?

7.2.2.3 Rôle déterminant des métiers dans la féminisation régionale

Au niveau d'une région, les métiers jouent un grand rôle dans la détermination de la participation féminine. Il s'agit alors ici de déterminer les métiers ou groupes de métiers qui amènent à considérer certaines régions comme positives ou négatives pour l'emploi féminin. En effet, l'importance relative de certains métiers dans une région peut induire, sur la base de la participation moyenne nationale, un niveau de participation féminine sur-représentée ou sous-représentée dans cette région. Cet élément de l'analyse est le reflet de l'effet de structure qui traduit la propension ou non de la région à se féminiser et qu'elle doit à sa composition sectorielle, suivant que les métiers fortement féminisés y sont largement représentés ou pas.

La forte participation féminine en Franche Comté est essentiellement imputable aux ouvriers et manoeuvres et les employés non qualifiés, ces deux métiers représentant sur l'ensemble des métiers retenus (nomenclature à 23 métiers) 38 % et 13 % respectivement de l'emploi salarié féminin dans la région. On peut associer également la féminisation de la région à la sur-représentation des femmes dans les métiers de mécanographe, opérateur, perforateur. Au total, on constate une forte représentativité des femmes dans les métiers non qualifiés qui induit une forte participation féminine en Franche Comté.

Quant au Centre, la forte participation féminine résulte d'une grosse proportion de femmes dans les métiers de faible qualification du secondaire et notamment d'une participation féminine très fortement supérieure à la moyenne nationale chez les ouvriers spécialisés et manoeuvres, métiers qui représentent dans la région environ 24 % de l'emploi féminin salarié si l'on considère la nomenclature à 23 métiers ou 16 % si l'on s'intéresse à l'ensemble des métiers.

En revanche, l'Île-de-France se distingue par une structure d'emploi nettement plus favorable à la participation féminine, ce d'autant plus que le tertiaire, où la part des femmes est en général largement supérieure à celle constatée sur le plan national, y est prédominant. Ainsi, on note une représentativité des femmes supérieure à la moyenne chez les employés administratifs qualifiés, les professeurs, professions littéraires et de l'information, les cadres administratifs supérieurs et les cadres moyens. L'évolution positive des effectifs féminins dans ces deux derniers métiers est l'illustration du fait que l'intégration croissante des femmes dans les métiers qualifiés, surtout les métiers du tertiaire, reste limitée aux grandes métropoles.

La Bourgogne a quant à elle une forte proportion des femmes supérieure à la moyenne dans les métiers du secondaire non qualifiés à savoir les ouvriers qualifiés verre, photographie et les ouvriers qualifiés du textile, cuirs et peaux, chimie.

A l'opposé, on distingue les régions à participation féminine relativement faible dans la plupart des métiers.

En Alsace où les ouvriers spécialisés et manoeuvres restent un métier très important, on note une forte sous-représentativité des femmes par rapport à la moyenne nationale. Cette sous-représentativité est également effective chez les ouvriers qualifiés de la radio-électricité, papier-carton et autres spécialités de l'alimentation, et chez les ouvriers qualifiés verre, photographie.

Le Nord voit sa participation féminine faible du fait de la sous-représentativité des femmes dans presque tous les métiers et surtout tertiaires, même s'il y a une relative propension pour elles à être présentes dans les métiers de la santé et les métiers non qualifiés du commerce et de la distribution.

Enfin, la Provence-Alpes-Côte d'Azur-Corse, compte tenu de la représentation des métiers dans l'emploi salarié de cette région, se caractérise par une sous-représentation des femmes par rapport à la moyenne nationale dans presque tous les métiers, à l'exception des professeurs, professions littéraires et de l'information, des ouvriers qualifiés du travail des étoffes.

L'analyse que nous venons de faire montre l'importance des facteurs économiques dans la structuration de la participation féminine régionale, et ce par le biais des métiers. Autrement dit, la féminisation des emplois dans une région est liée à l'importance relative d'un certain nombre de métiers, lesquels traduisent la structure des emplois offerts.

D'autre part, il a été constaté que plus la part des métiers non qualifiés surtout les métiers du secondaire - était élevée dans une région (respectivement faible), plus celle-ci était féminisée (respectivement faiblement féminisée). On peut alors supposer, ainsi que le prouvent d'ailleurs certains travaux de l'INSEE [18], que l'emploi salarié féminin, contrairement aux tendances observées sur le plan national est lié à l'industrialisation de la région et non à sa tertiarisation. Certes les cas du Nord-Pas-de-Calais et de la Lorraine font exception à la règle, puisque malgré l'existence d'une forte proportion d'ouvriers dans ces régions, on enregistre une forte sous-activité des femmes. Peut-être cette liaison négative est-elle due au poids des charges familiales (la fécondité y est très forte) ou encore aux tensions sur le marché du travail (la crise a durablement frappé les deux régions et ne favorise pas la multiplication des opportunités d'emplois que bénéficierait une main-d'œuvre potentiellement latente dont une grande partie serait constituée des femmes qui pour l'instant se déclareraient inactives) ?

Par ailleurs, la féminisation d'une région dépend de la répartition par sexe des emplois dans les métiers du secondaire. Plus la part des femmes dans ces métiers est importante, plus la féminisation d'une région est importante également. Ce schéma causal n'existe pas avec les métiers du tertiaire et on pourrait expliquer cela par la diffusion presque homogène du tertiaire à travers les régions.

Dans les régions Provence-Alpes-Côte d'Azur, Languedoc-Roussillon, aucun facteur structurel n'est à même d'expliquer leur sous-activité et il faudrait trouver des éléments d'explication soit dans les traditions historiques, soit dans le caractère particulièrement stagnant du marché de l'emploi, l'ampleur du chômage étant assez manifeste dans ces régions depuis la crise.

En somme, on peut dire que la diversité de la participation féminine régionale est à lier à l'existence plus ou moins importante des métiers fortement féminisés. Certes, on pourra aussi évoquer d'autres facteurs qui interviennent marginalement tels que l'urbanisation. En effet, la diversité des emplois offerts est d'autant plus grande que l'agglomération urbaine est importante. De plus, l'activité des salariées est très faible dans les communes rurales, et ce dans la mesure où la rareté de ces emplois est perceptible à proximité de leur domicile.

Sur le marché du travail, certains modèles théoriques - par exemple le modèle de marché dualiste - ont tendance à regrouper les femmes et les jeunes - ce sont des catégories dites infériorisées ou secondaires -, du fait d'une certaine similitude de leurs caractéristiques

comportementales. En quittant le présent paragraphe consacré à la participation féminine régionale, nous allons nous tourner vers le suivant dans lequel nous analyserons le comportement des salariés suivant leurs classes d'âge et par métier d'une part, et par région d'autre part. Cette analyse nous permettra de vérifier si le regroupement généralement retenus des jeunes et des femmes est pertinent ou pas.

7.3 ANALYSE DE LA STRUCTURATION DES METIERS ET REGIONS PAR LE FACTEUR AGE

Depuis la dernière décennie, le marché de travail connaît constamment de nombreuses tensions dont l'intensité et l'amplification reflètent la crise de l'emploi que traverse la France. Autrement dit, les difficultés de l'accès à l'emploi sont grandes et l'une des manifestations les plus perceptibles est la montée non contenue du chômage qui frappe différemment les catégories de main-d'oeuvre repérées selon leur âge. C'est ainsi qu'en 1980, les taux de chômage enregistrés pour les classes d'âge 18 - 24 ans, 25 - 39 ans, 40 - 49 ans, 50 - 59 ans sont respectivement de 36,6 %, 32,4 %, 11,6 % et 13,2 % [19]. Les différences que révèlent ces taux témoignent en quelque sorte le rôle discriminant de l'âge sur l'offre de travail. Cette discrimination sera surtout appréhendée à travers les métiers d'une part et les régions d'autre part.

L'accent dans cette analyse sera mis sur une classe particulière où les différences sont importantes aussi bien à travers les métiers qu'au niveau des régions : les jeunes, c'est-à-dire les salariés de moins de 25 ans, qui n'ont que la faveur de certains métiers où la mobilité est très forte, celle-ci étant le reflet de la politique de rationalisation de la main-d'œuvre mise en place par les entreprises et qui est favorisée à tort ou à raison par les politiques catégorielles d'emploi des pouvoirs publics.

7.3.1 L'âge comme facteur de structuration des salariés suivant les métiers

L'analyse des effets croisés entre profession et âge indique l'existence de fortes différences au niveau des métiers dans les classes d'âge moins 25 ans et 55 ans et plus. Ces différences, moins importantes pour les âges moyens traduisent en filigrane le type des métiers qu'investissent les jeunes à un degré moindre les plus de 55 ans. Compte tenu de la

faible représentativité relative de cette dernière classe dans l'échantillon et du fait que les salariés ici ne sont plus qu'à quelques pas de leur sortie d'activité, nous n'insisterons pas assez souvent sur les différences constatées dans cette classe d'âge.

Les salariés de moins de 25 ans sont assez représentés dans les métiers peu qualifiés du tertiaire en général. Leur proportion est forte dans les métiers non qualifiés du commerce et distribution, les métiers non qualifiés de l'hôtellerie, chez les ouvriers spécialisés et manoeuvres, ces trois métiers où ils s'opposent à leurs aînés, relativement sous-représentés. D'autre part, la sur-représentativité des moins 25 ans et des 25 - 31 ans s'oppose à la sous-représentativité des autres classes d'âge chez le personnel qualifiés des techniques comptables, les employés non qualifiés, les ouvriers qualifiés de la radio-électricité, du papier-carton et autres spécialités de l'alimentation ; les mécanographes, opérateurs, perforateurs peuvent également faire partie de ce groupe bien que l'on note une représentativité relativement moyenne de la classe d'âge 32 - 39 ans.

Cette opposition reste également valable pour certains métiers où le niveau de qualification est assez élevé notamment les cadres moyens, les cadres administratifs supérieurs, les métiers techniques de l'hôtellerie, les métiers techniques de la distribution, mais qui à l'inverse du groupe précédent, traduisent une sous-représentativité des moins 25 ans, voire des 25 - 31 ans et une sur-représentativité de leurs aînés. Ces derniers métiers sont ceux où l'entrée se fait à un âge moyen (cadres moyens, métiers techniques de la distribution, métiers techniques de l'hôtellerie) ou aux âges élevés (cadres administratifs supérieurs). En effet, le niveau de qualification requis pour les métiers techniques de la distribution par exemple nécessite une durée d'étude assez longue, les jeunes n'entrant dans ces métiers que vers 25 ans, ce qui justifie leur très forte sous-représentativité. Toutefois, il y a une représentativité relativement moyenne des salariés de 25 - 31 ans et au-delà le niveau de représentativité est presque le même pour les autres classes d'âge, ce qui semble montrer qu'au-delà de 32 ans, beaucoup de salariés ne quittent pas ces métiers.

Les cadres administratifs supérieurs ont des caractéristiques proches de celles des métiers techniques de la distribution avec une entrée tardive des jeunes et qui par rapport aux autres métiers est longue. C'est ici que se manifeste la plus forte sous-représentativité des jeunes. D'ailleurs, dans ce métier, la représentativité est positivement liée à l'âge et la très forte représentativité des plus de 55 ans indique le comportement des salariés de ces métiers qui y restent longtemps actifs et par conséquent, les quittent au-delà de l'âge de la retraite.

Tableau 24 : Evolution de l'emploi salarié par classe d'âge entre 1976 et 1981

Professions ↓	Ages→	- 25	25-31	32-39	40-47	48-54	55 +
OQ trav. étoffes				+	-	+	-
Métiers de la santé		+	-	-	+		++
PQ paramed +soc.		--	-	0	+	0	+
M soins personnels		++	--	+	-	-	-
PQ tech. comptables		+	++	+	-	-	-
Méc., opérat., perfor.		-	-	+	+		
M. nonQ comm+dis		++	-	--	+		
Emplois non qualifiés		-	+	+	+	-	-
M. tech. hôtellerie		++	-	+	+	+	--
MQ vente hôtellerie		++	-	0	-	-	0
Cadres moyens		-	-	++	+	--	+
Comptables, techn. banques		+	-	+	-	+	
M. nonQ hôtellerie		++	++	--	-	--	++
Professeurs, litt. Inform.		-	-	0	++	0	++
Emplois adm. qualifiés		--	0	+	+	-	
OQ text., cuirs, chimie		+	0	-		+	+
OS + manœuvres		+	0	-	0	+	0
OQ radio, élec, pcar		+	++		--	-	--
M. tech. distrib.		+	--		+	+	0
Cadres adm. Sup		-	-	0	0	-	-
OQ verre photo		-	-	-	+		--
Tech. et dessinat.		0	0	+		-	0
Autres OQ		0	0	0	0	+	0

++ : croissance très forte

+ : croissance forte

0 : stabilité

- : décroissance

-- : décroissance très forte

La représentativité des autres classes d'âge intermédiaires peut se traduire par les effets de promotion interne.

Les métiers qualifiés que nous venons d'analyser sont ceux où la représentativité des femmes est relativement moyenne ou faible (cf. 7.2.1). Toutefois les conclusions que nous venons d'en tirer ne sont pas propres à tous les métiers qualifiés et on peut d'ailleurs s'en apercevoir dans la structuration du personnel qualifié des soins para-médicaux et sociaux qui est un métier à forte participation féminine. Il s'agit d'un métier dont la qualification exige aussi de longues années d'études, et les jeunes - les femmes en majorité - y entrent tardivement. On note plutôt une forte représentativité des 48 ans et plus qui semble liée aux départs précoces dans ce métier.

La proportion des jeunes est également faible chez les ouvriers qualifiés du travail des étoffes, métiers où l'entrée est fermée aux jeunes malgré son bas niveau de qualification. Ce sont surtout les classes d'âge élevées qui sont très présentes dans ce métier et qui pourraient se justifier par une sorte de promotion interne permettant le passage des salariés du statut de manoeuvre à celui d'ouvrier qualifié, passage qui se détermine par l'expérience qui n'est pas l'apanage des jeunes.

De cette analyse, on peut tirer quelques conclusions partielles :

- les jeunes sont en général très représentés dans les métiers peu qualifiés du tertiaire ;
- leurs aînés sont très représentés dans les métiers qualifiés même si on les trouve moins représentés aux âges élevés dans certains métiers qualifiés tels que personnel qualifié des soins para-médicaux et sociaux, professeurs, professions littéraires et de l'information, où malgré la sélection à l'entrée des moins de 25 ans, on retrouve une forte proportion des salariés de 25 - 40 ans et qu'on peut globalement caractériser comme des métiers à prédominance "jeunes".

Ces conclusions nous amènent à tracer le lien avec le niveau de formation des salariés de moins de 25 ans et de constater qu'ils sont d'autant plus représentés dans un métier que le niveau de formation requis est faible, même si cette variable n'apparaît pas dans les données, ce qui aurait permis une meilleure appréciation. D'autre part, les caractéristiques des jeunes dans les métiers sont très similaires à celles qu'on a déjà retrouvées chez les femmes en général (cf. 7.2.1).

Cette situation n'est pas assez surprenante dans la mesure où deux métiers sur trois ont une part relativement supérieure à la moyenne de femmes de moins 25 ans.

Cette part est relativement importante dans les métiers où les jeunes sont soit fortement, soit moyennement représentés tels que métiers des soins personnels, mécanographes, opérateurs, ouvriers qualifiés des textiles, cuirs et peaux, chimie, ouvriers qualifiés du travail des étoffes, personnel qualifié des soins para-médicaux et sociaux. On remarquera au passage qu'il s'agit des métiers aux valeurs féminines. Les jeunes femmes sont également très présentes dans les métiers qualifiés où les moins 25 ans sont en général sous-représentés : métiers techniques de l'hôtellerie, comptables et techniques des banques, cadres administratifs supérieurs. Dans ces métiers, on ne trouve pas assez de femmes aux âges élevés et d'autre part ils se caractérisent par une entrée tardive des jeunes. Le niveau de formation des jeunes femmes étant plus élevé que celui des jeunes gens, on peut penser que les premières investissent plus rapidement que les seconds ces métiers, ce qui se traduit par la sous-représentativité des derniers.

En somme, ces résultats complètent parfaitement ceux que nous avons eus en analysant la participation féminine et où il ressortait qu'il y avait une polarisation des femmes dans les métiers peu qualifiés du tertiaire. De plus, la majorité de ces femmes ont moins de 25 ans, c'est-à-dire qu'elles sont à une période très importante comme tous les autres jeunes - de leur insertion professionnelle. Etant données la précarisation de l'emploi qu'offrent ces métiers et les difficultés d'insertion que connaissent les jeunes (leur taux de chômage est de 28% en 1981), on est en droit de relativiser le développement continu de l'emploi féminin puisque les femmes accèdent généralement aux emplois instables d'une part et aussi d'analyser le rôle que jouent désormais les jeunes sur le marché du travail d'autre part, ce à la lumière d'une panoplie de mesures qui sont prises, en leur faveur dans le cadre des politiques spécifiques d'emploi.

7.3.2 Main-d'oeuvre juvénile et son rôle spécifique : sur le marché de travail

L'analyse précédente vient de mettre à nu les métiers où les jeunes excellent. Il s'agit comme on l'a vu des métiers où le niveau de formation initiale est faible et où le renouvellement de la main-d'oeuvre est permanent. Ces métiers joueraient un rôle central dans l'insertion des jeunes puisqu'ils accueillent une part croissante de ceux-ci.

Tableau 25 : Discrimination profession-âge selon le sexe

Professions ↓	Ages→	- 25	25-31	32-39	40-47	48-54	55 +
OQ trav. étoffes		++				++	-
Métiers de la santé		-	+	+	+	--	
PQ paramed +soc.		-	-	++	+	--	+
M soins personnels		--	--	++	-	+	++
PQ tech. comptables		0	-	+	--	+	
Méc., opérat., perfor.				-	-	++	
M. nonQ comm+dis		--	-	+	+	+	0
Emplois non qualifiés					-	-	+
M. tech. hôtellerie			+	0	0	0	0
MQ vente hôtellerie		+	-	+			
Cadres moyens			0	0			+
Comptables, techn. banques		-	0			0	-
M. nonQ hôtellerie		+			-		
Professeurs, litt. Inform.			0	0	-	+	0
Emplois adm. qualifiés		-	+	0		-	
OQ text., cuirs, chimie		-			-	+	+
OS + manœuvres		+		-	+	-	0
OQ radio, élec, pcar		--	0		-	-	++
M. tech. distrib.		++	0				0
Cadres adm. Sup		--	-	-	+	+	+
OQ verre photo		0	++		+	-	
Tech. et dessinat.		+	0		-	+	
Autres OQ		-	-			+	+

++ : croissance très forte

+ : croissance forte

0 : stabilité

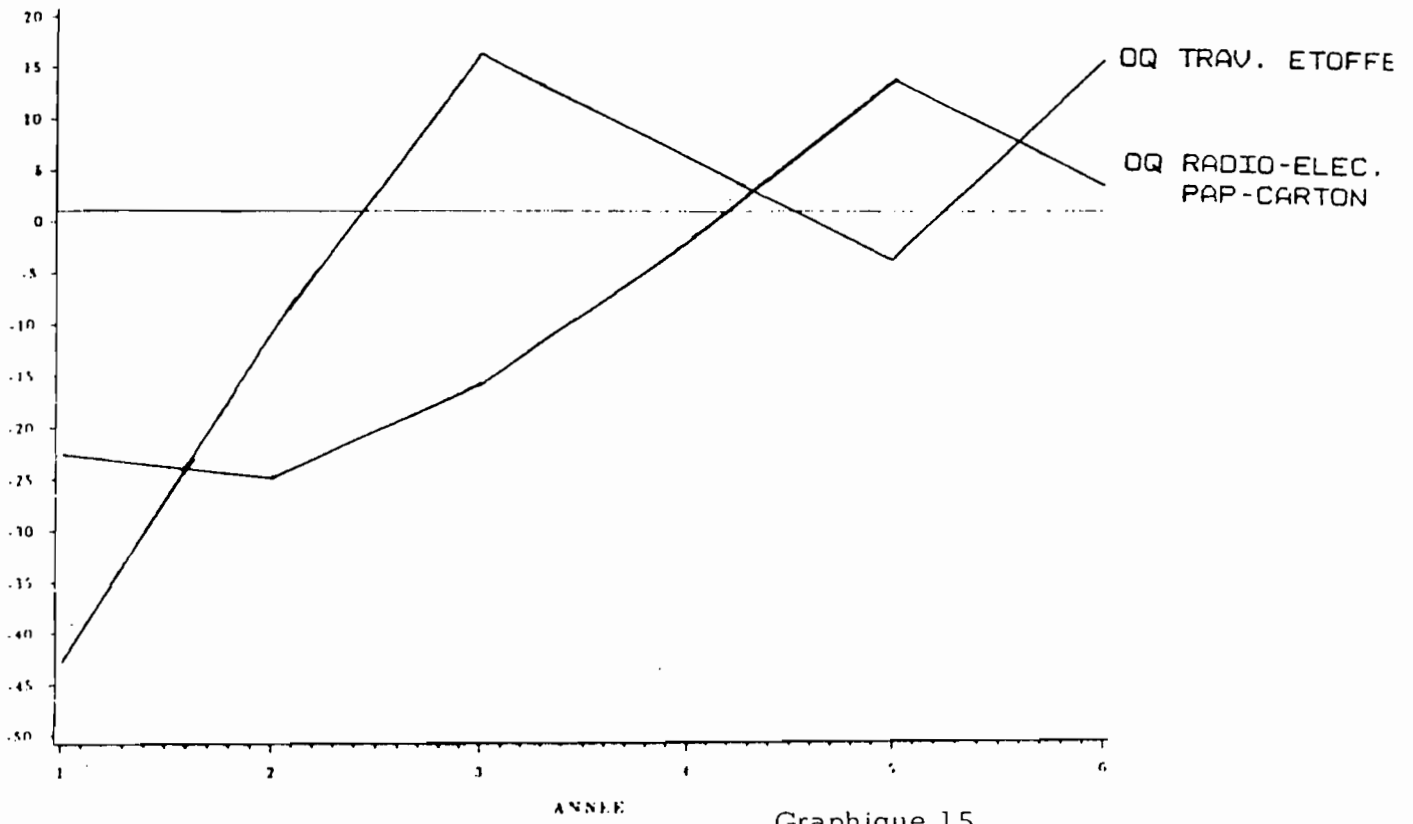
- : décroissance

-- : décroissance très forte

D'ailleurs, sur la période 1976-1981, les évolutions les plus significatives sont à l'actif des métiers des soins personnels métiers non qualifiés du commerce et distribution, métiers non qualifiés de l'hôtellerie (cf. tableau 24) où les effectifs des moins 25 ans augmentent régulièrement. Ces évolutions sont moins fortes dans les métiers du secondaire.

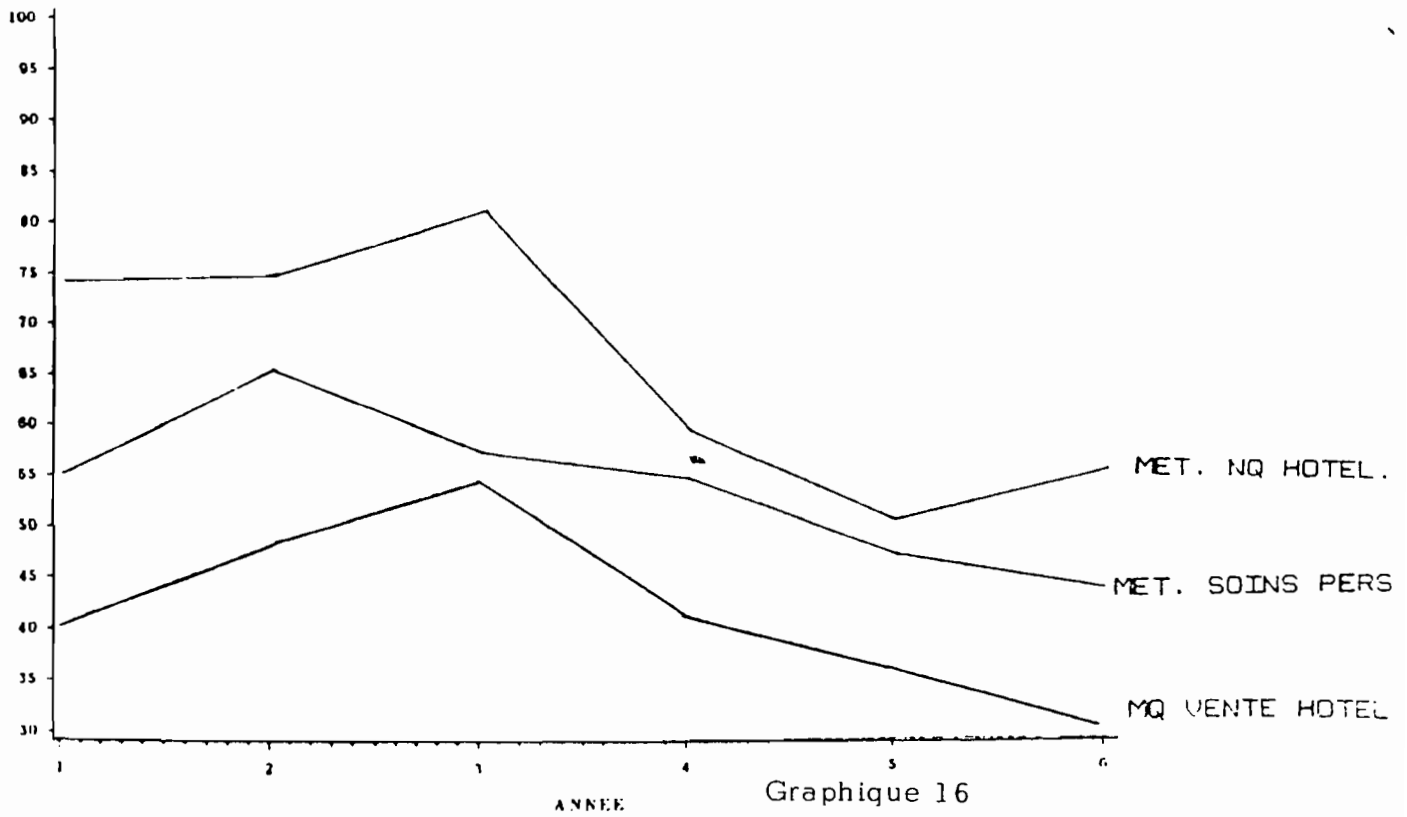
Les éléments de l'analyse que nous possédons jusque maintenant ne nous donnent qu'un aperçu partiel quant au comportement des jeunes sur le marché du travail. Certes, ils occupent des emplois peu qualifiés, mais ces résultats sont insuffisants pour juger leur statut précaire et qui se traduit par une mobilité sur le marché de travail. Ainsi, en 1981, on dénombre un flux minimum d'embauches de 852.000 jeunes de moins 25 ans parmi lesquels 586.000 étaient étudiants, chômeurs, inactifs ou militaires du contingent un an auparavant et 266.000 qui ne sont plus dans la même entreprise qu'un an auparavant [20]. Ce flux est minimum, car beaucoup parmi ces jeunes ont changé plusieurs fois d'emploi au cours de l'année, soit à cause d'un licenciement, soit à cause de la nature de leur contrat qui est à durée déterminée ou d'intérim. On touche ici le nœud du problème de la précarisation de l'emploi des jeunes car ces formes d'emploi s'appliquent souvent à eux : 45% des jeunes de moins 25 ans ont bénéficié des contrats à durée déterminée en 1985 [21] et l'augmentation de ces contrats est de 23% entre 1984 et 1985. L'ampleur du phénomène est significative pour les femmes jeunes. En effet, l'enquête complémentaire à l'enquête emploi 1980 révèle que 10,4% des salariés de moins 25 ans occupent à cette époque un emploi intérimaire ou sous contrat à durée déterminée contre 3,8% pour l'ensemble des femmes salariées et 2,7% pour l'ensemble des hommes salariés [22]. Toujours selon la même enquête, 24% de femmes de moins 25 ans contre 10% pour l'ensemble des femmes embauchées depuis moins d'un an avaient trouvé un emploi intérimaire ou à durée déterminée. Ce taux n'est que 17% pour les jeunes gens. Cette précarisation accrue de l'emploi des jeunes se fait généralement dans les métiers où ils sont très représentés et que nous avons vu plus haut, métiers qu'ils exercent assez souvent dans des entreprises de petite taille qui offrent à certains d'entre eux sans formation des statuts intermédiaires entre emploi et formation (stage, apprentissage, contrats emploi-formation etc.) qui leur assurent une première expérience professionnelle, le temps d'espérer trouver ailleurs un emploi stable aux conditions meilleures, ce qui rappelle, concernant l'emploi des jeunes, la distinction faite par VINCENS entre emplois d'attente et emplois définitifs [23]. Ces statuts intermédiaires s'inscrivent dans le cadre des mesures catégorielles prises pour l'incitation à l'insertion des jeunes, mais elles semblent constituer un parcours à obstacles par lesquels un

EVOLUTION DE L'EMPLOI PAR PROFESS. EN PACC (1976-1981)



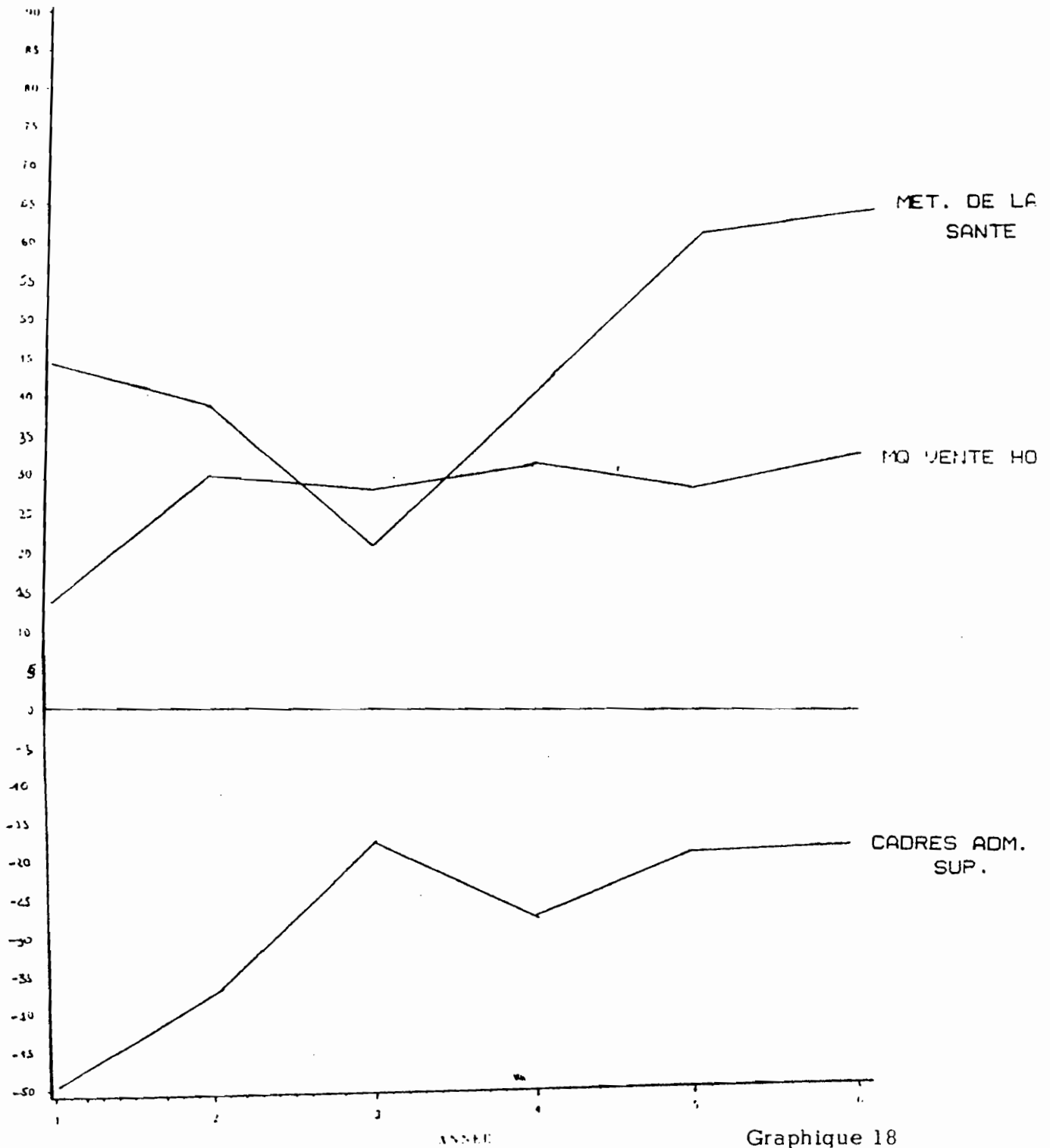
Graphique 15

EVOLUTION DE L'EMPLOI PAR PROFESS. EN PACC (1976-1981)



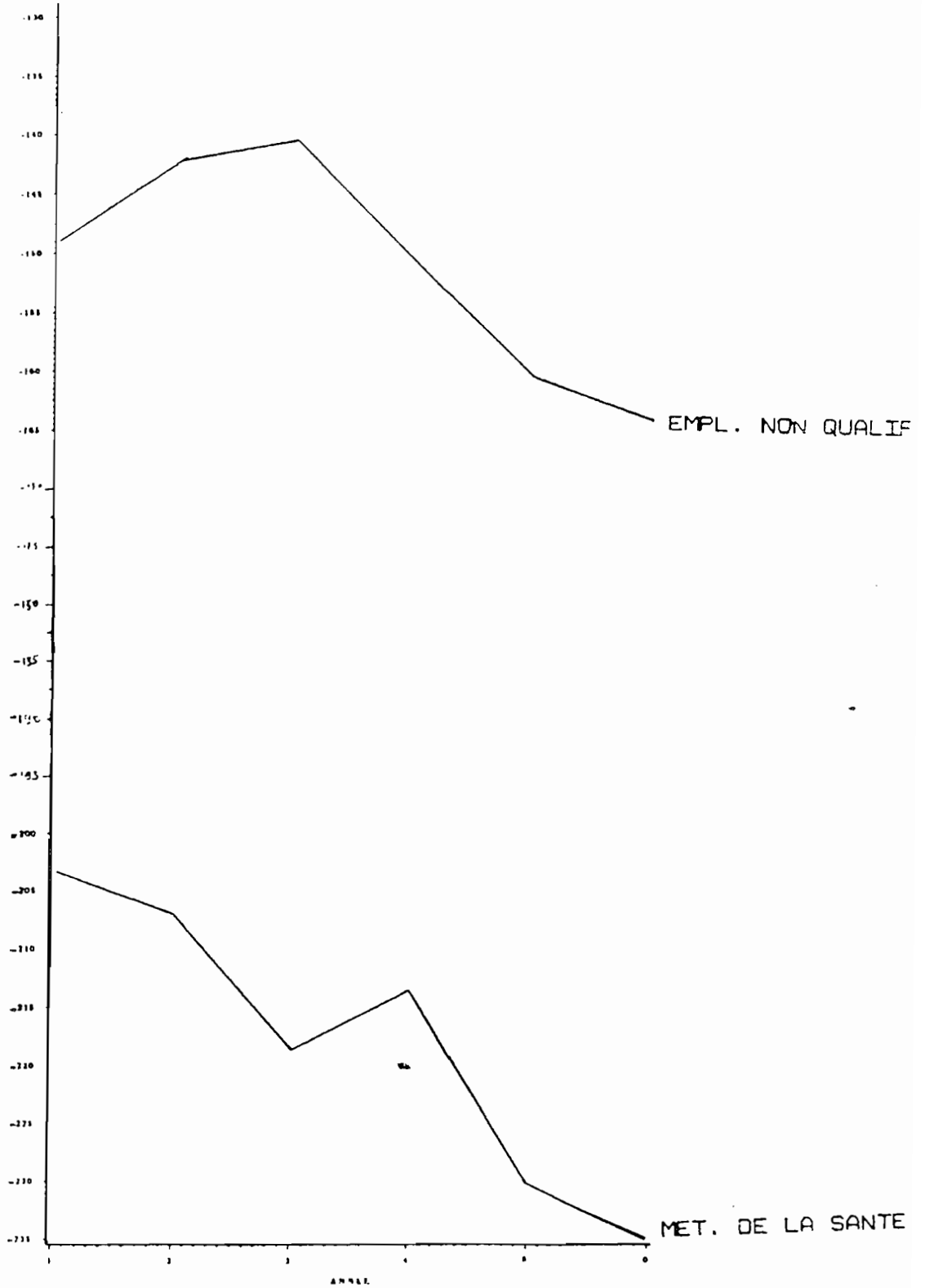
Graphique 16

EVOLUTION DE L'EMPLOI PAR PROFESS. EN BRE (1976-1981)



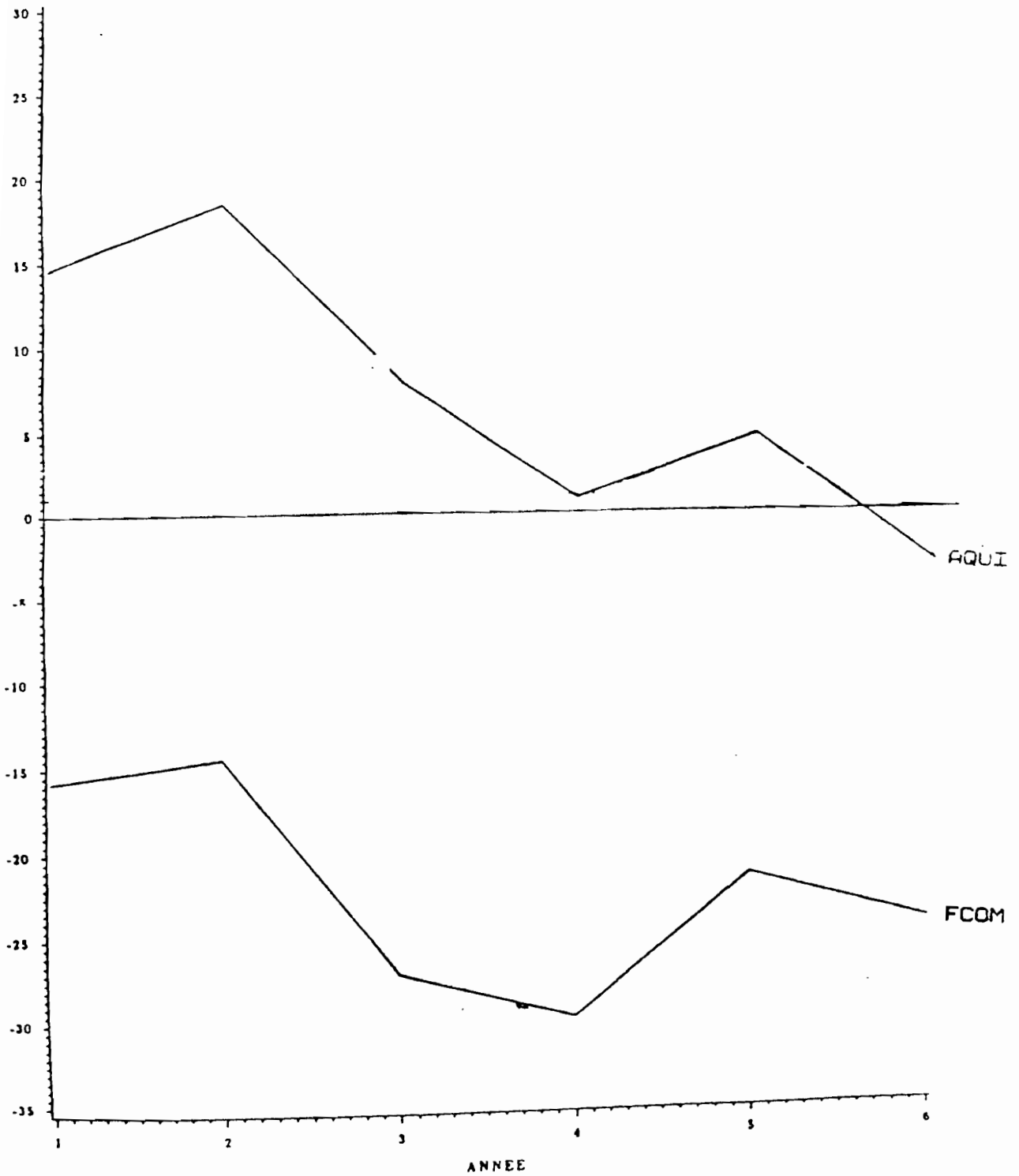
Graphique 18

EVOLUTION FEMINISATION RELATIVE DES PROFESSIONS (1976-1981)

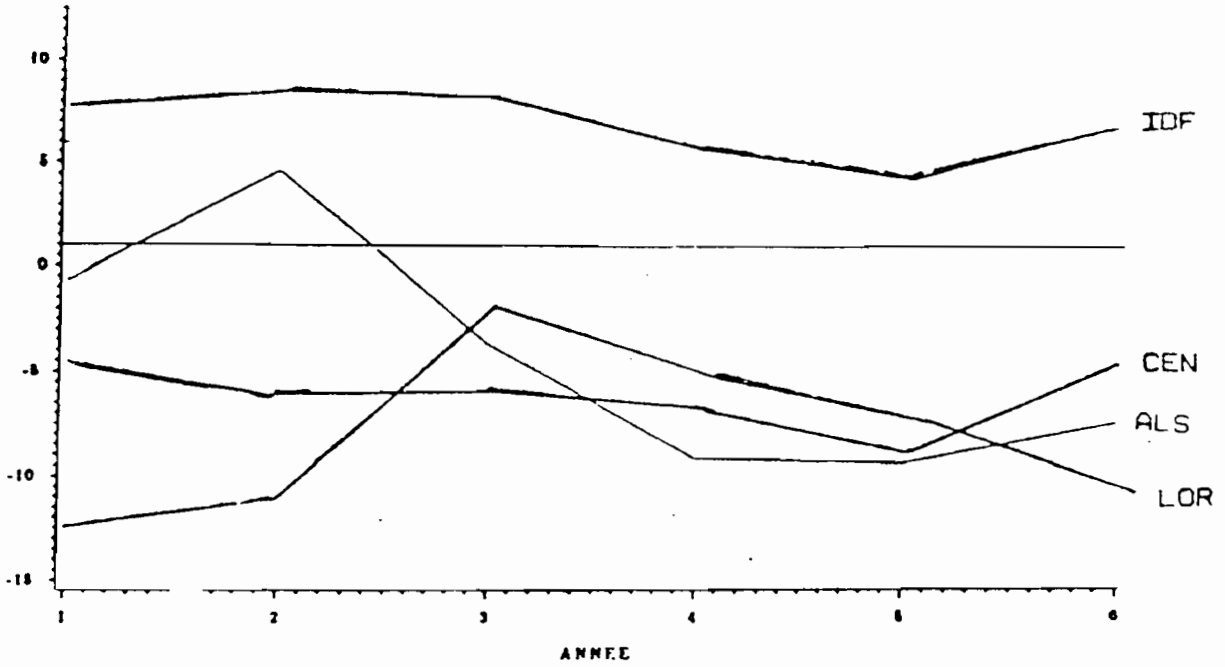


Graphique 19

EVOLUTION FEMINISATION RELATIVE REGIONALE (1976-1981)

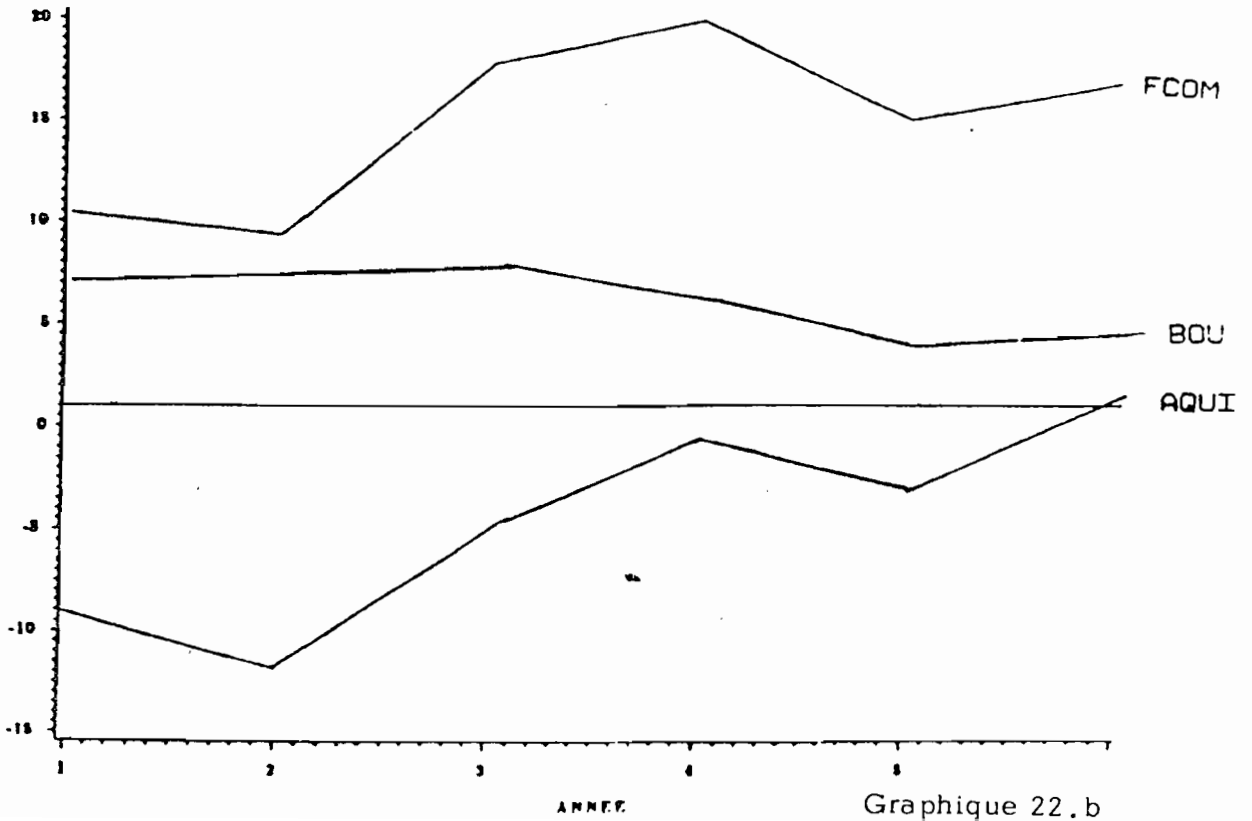


EVOLUTION FEMINISATION ABSOLUE REGIONALE (1976-1981)



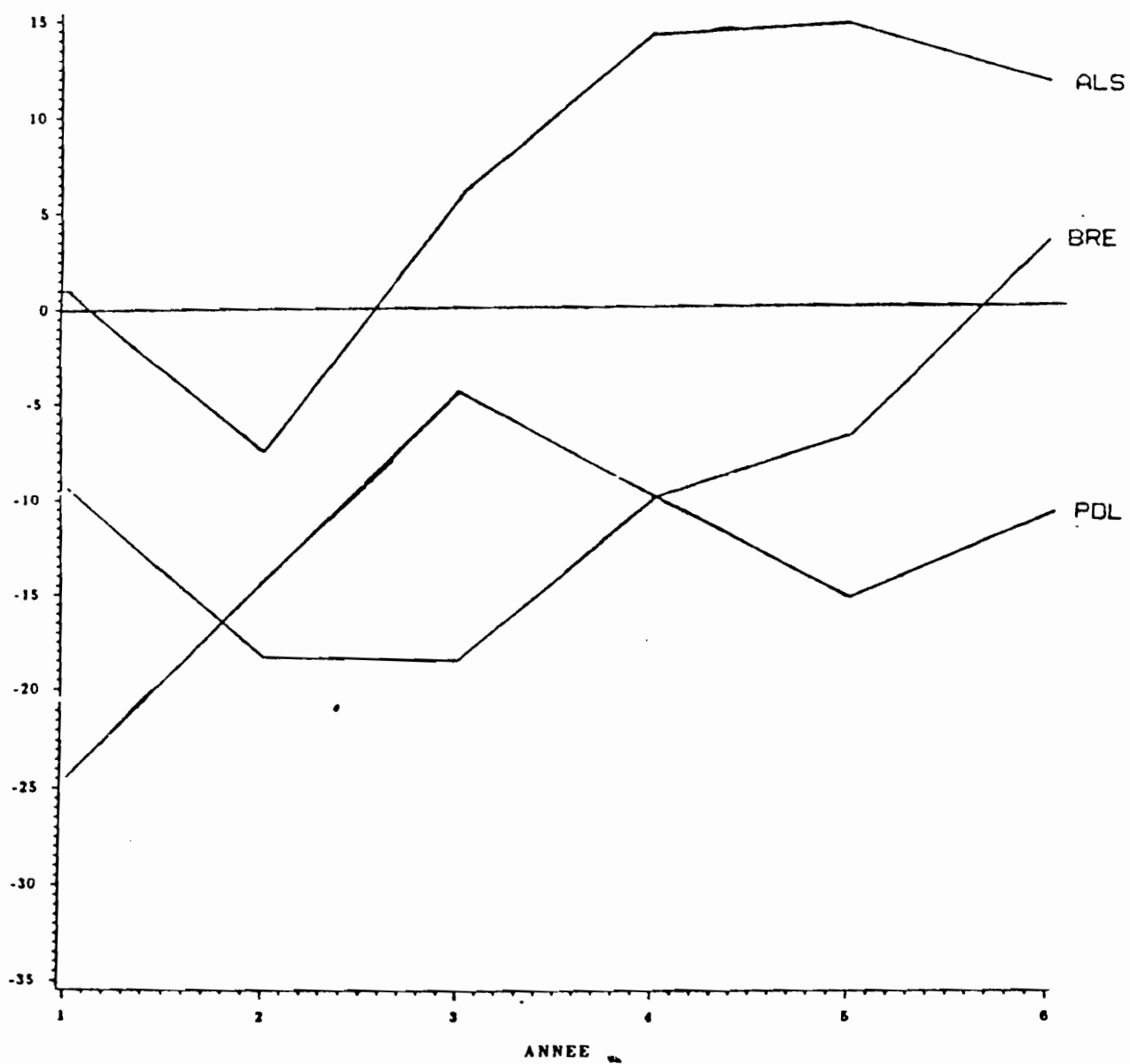
Graphique 22.a

EVOLUTION FEMINISATION ABSOLUE REGIONALE (1976-1981)



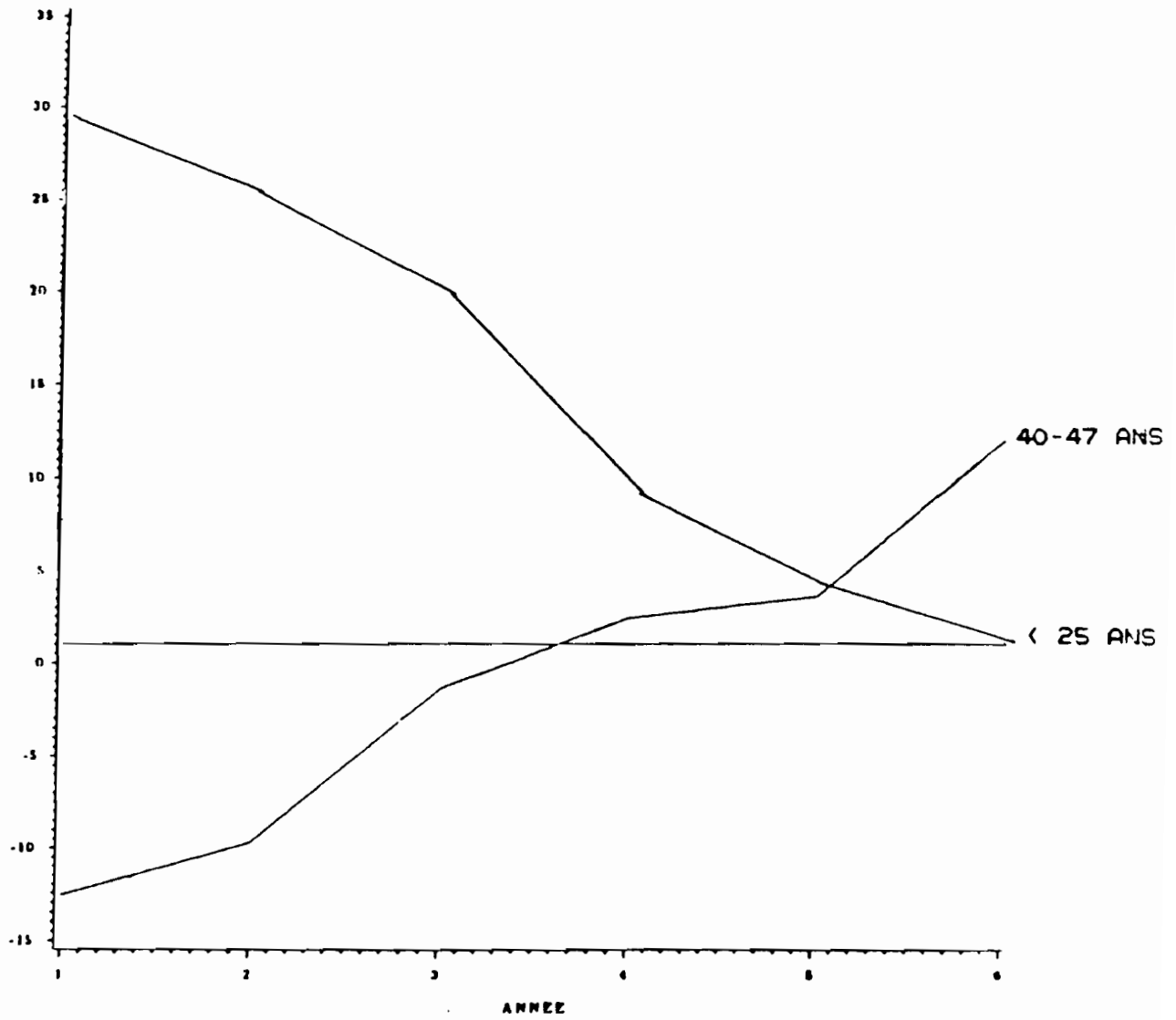
Graphique 22.b

EVOLUTION DEFEMINISATION RELATIVE REGIONALE (1976-1981)



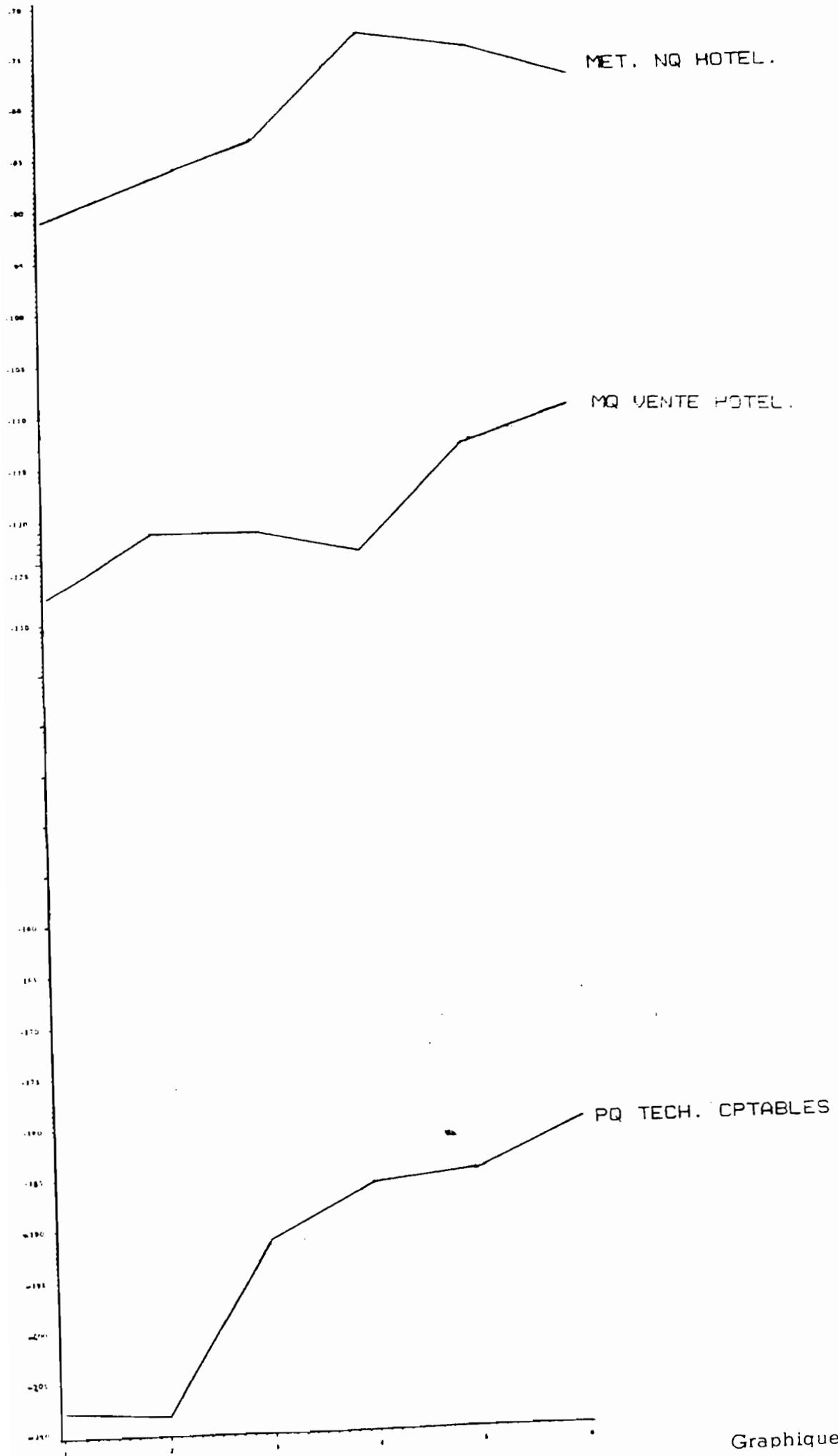
Graphique 24

EVOLUTION DE L'EMPLOI PAR AGE EN ALSACE



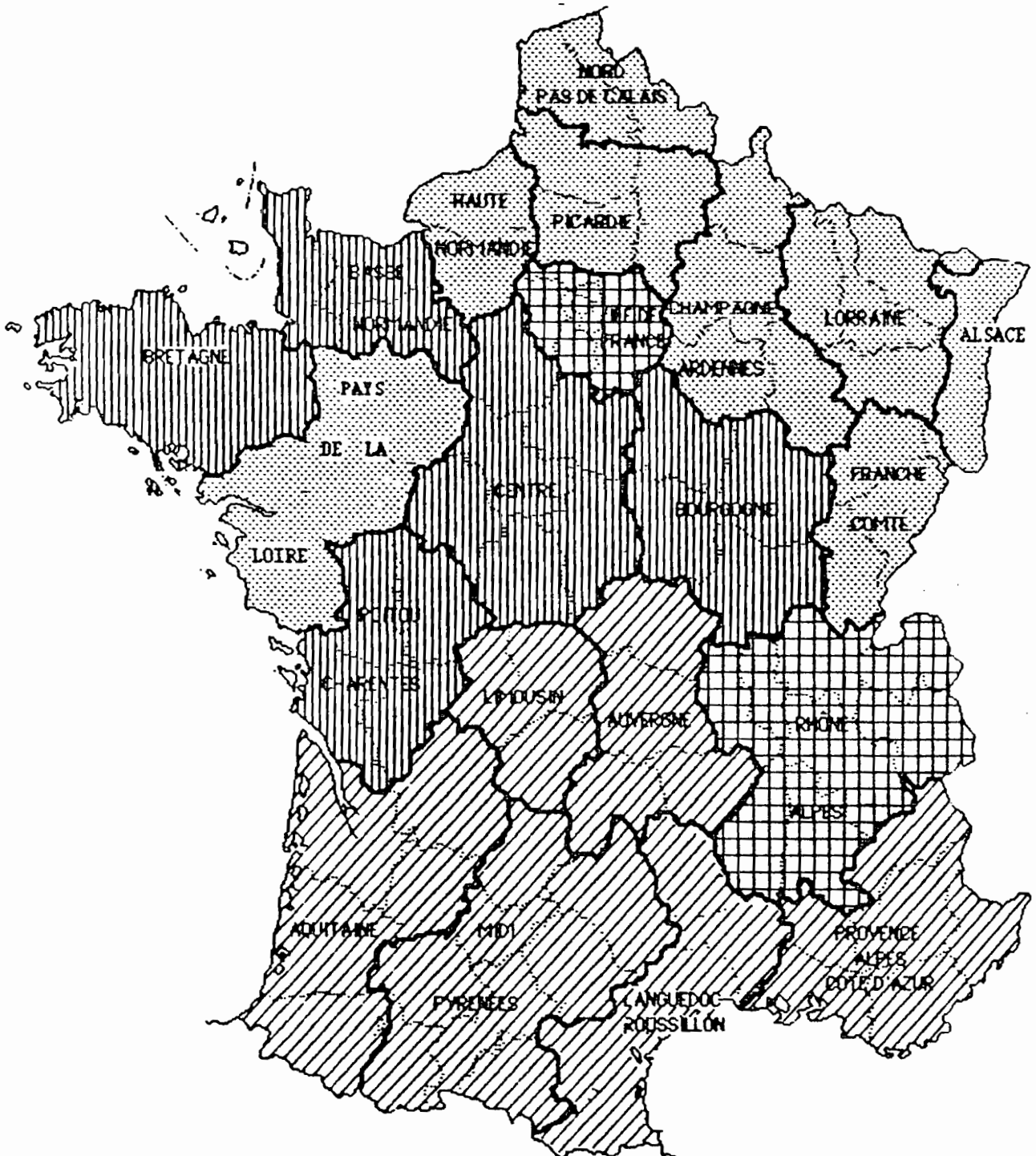
Graphique 25

EVOLUTION DEFEMINISATION RELATIVE DES PROFESS. (1976-1981)

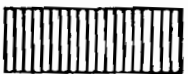


Graphique 20

TYPOLOGIE PAR AGE DES REGIONS



Sur-représentativité aux âges intermédiaires (25-54 ans)
et sous-représentativité aux âges extrêmes.



Sur-représentativité des classes d'âge extrêmes et sous-représentativité
des classes d'âges intermédiaires.



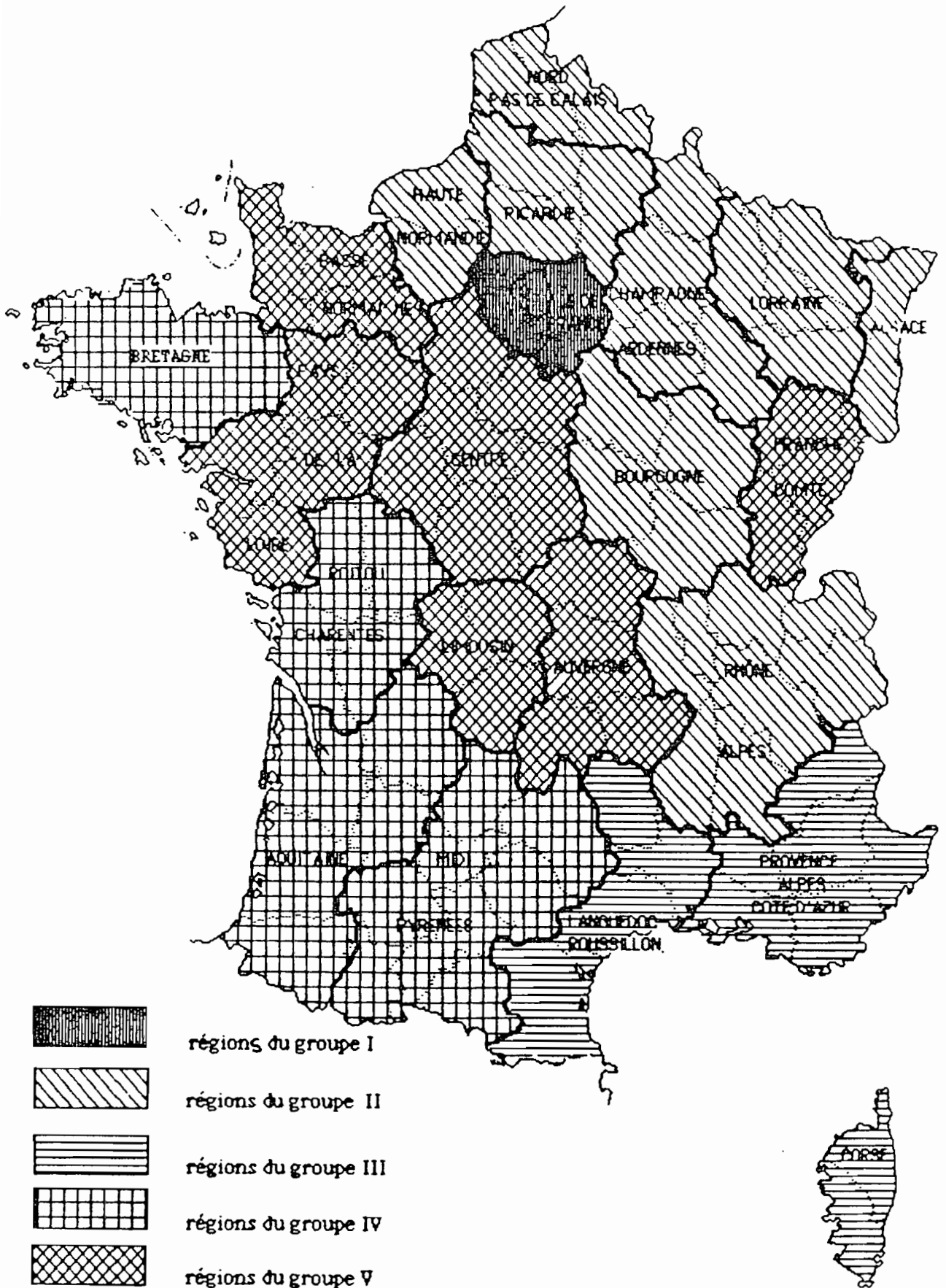
Sur-représentativité des jeunes et sous-représentativité des autres
classes d'âges.



Sous-représentativité des jeunes et sur-représentativité des autres
autres classes d'âges.



TYPLOGIE REGIONALE DES METIERS



CONCLUSION GENERALE

A l'origine, le modèle Log-linéaire a été utilisé pour la détection, dans un tableau de petite taille, des liens existant entre deux ou plusieurs variables, dans la tradition de l'analyse de la variance. Du fait de ses prolongements analytiques à travers son approche informationnelle, il a désormais le statut d'un véritable outil statistique, adapté à l'étude exploratoire des données économiques se présentant sous la forme d'un tableau de contingence multidimensionnel. Peu ou prou, ce modèle a permis sur la période 1976-1981, de dégager quelques tendances lourdes et des effets de structure ayant modulé l'emploi salarié en France. Ainsi, a-t-on pu observer des évolutions significatives, notamment :

- une diminution de l'emploi salarié assez marquée dans les métiers du secondaire. Bien que cette diminution touche les ouvriers dans l'ensemble, son ampleur est beaucoup plus importante chez les ouvriers non qualifiés (les ouvriers spécialisés et manœuvres perdent 620.000 emplois entre 1975 et 1982) que chez les ouvriers qualifiés, parmi lesquels les ouvriers qualifiés du textile, cuirs et peaux, chimie. Cette baisse, qui néanmoins s'inscrit dans une perspective de long terme, s'est amplifiée sur la période étudiée, faisant passer la part des ouvriers de 38 % à 35 %. Ce mouvement est particulièrement imputable au déclin de l'industrie, et aux multiples restructurations qui y ont été opérées, afin de juguler la crise, dont les incidences sont négatives sur le niveau de l'emploi. Les professions ouvrières devraient encore subir davantage les contre coups de la crise dans les années à venir, où les perspectives de créations d'emplois ne sont pas fastes. Il y a même une forte propension à la suppression d'emploi, et la DATAR a prévu une perte minimum de 100.000 emplois dans l'industrie pour l'année 1987 ; cette baisse affecterait principalement les ouvriers ;
- une croissance de l'emploi salarié dans le tertiaire qui représente 56,3 % de l'emploi national en 1981. Cette évolution est à lier à l'augmentation des effectifs dans les métiers de la santé (40 % des créations d'emplois entre 1975 et 1982) qui contribuent de façon déterminante à la progression des effectifs. On enregistre également une croissance de l'emploi chez les professeurs, professions littéraires et de l'information, le personnel qualifié des soins para-médicaux et sociaux, les cadres moyens, les cadres administratifs supérieurs, etc.. Pour les trois premiers métiers, on peut y voir le rôle prépondérant et décisif des pouvoirs publics, dans la mesure où les personnels de la santé et de l'enseignement relèvent généralement de statuts publics, qui influent de façon nette sur les modalités de la gestion de la main-d'œuvre et par conséquent sur l'emploi tertiaire. Il y a un tassement des effectifs dans les métiers de la santé en fin de période. Ils se présentaient auparavant comme des principaux pourvoyeurs de l'emploi salarié dans le tertiaire, et les difficultés actuelles de financement de

la sécurité sociale incitent à la rationalisation des prestations de la santé; l'effet manifeste y étant un frein des embauches ;

- une relative stabilité de l'emploi salarié dans les classes d'âge 32 ans et plus et une évolution chez les moins 25 ans qui s'oppose à celle des 25-31 ans : les emplois des moins de 25 ans connaissent une baisse régulière sur la période, nonobstant toutes les mesures incitatives prises dans le cadre de l'action des pouvoirs publics pour les préserver. Ils subissent de façon différenciée les effets de la situation dégradée de l'emploi, à cause de l'instabilité fréquente de leurs premiers emplois, favorisée par le développement des emplois précaires (intérim, contrats à durée déterminée, etc.), où ils se réfugient transitoirement en attendant les meilleures perspectives. A la fin de 1982 et au début de 1983, les embauches liées aux contrats de solidarité avaient apporté une amélioration passagère et depuis 1984, la croissance du chômage des jeunes reste vive. A l'inverse, la classe d'âge 25-31 ans enregistre plutôt une progression des effectifs, et l'on peut y voir l'effet de l'affermissement des qualifications, du fait de la transition par les « emplois d'attente » ou de la prolongation de la scolarité.

- une croissance relative de l'emploi régional, toutefois moins importante que celle observée dans les métiers. La croissance la plus forte est à mettre à l'actif de la région Languedoc-Roussillon, qui sur la période, se singularise par un taux de création d'emplois assez important (10 %). Cependant, il s'agit d'une région où l'évolution de l'emploi et celle de la population active sont ambivalentes ; elle figure parmi les régions qui gardent le peloton de tête en matière de chômage sur les vingt dernières années. D'autres régions telles que le Rhône-Alpes, l'Alsace, la Picardie et la Franche-Comté ont, en matière d'emploi, des évolutions assez favorables. Ce résultat semble surprenant en Franche Comté où sa texture économique, quasiment mono-industrielle, laisserait préjuger de la baisse de l'emploi, compte tenu du fait que la crise frappe surtout le secondaire. Peut-être ce résultat tient-il de la composante structurelle de la région ?

Ces premiers résultats ne sont pas en soi-même révolutionnaires. Tout au moins, ils mettent en lumière la robustesse de l'outil d'analyse et son adéquation pour la description des phénomènes économiques. Son originalité, c'est qu'il permet de révéler l'effet propre à chaque variable, c'est-à-dire l'effet pur, épuré de l'effet dû aux autres variables.

Certes, l'emploi est différencié selon l'âge, le sexe, la profession et la région. Mais cette différenciation se fait par la médiation de ses structures. Il apparaît alors que l'allocation du travail réalisée sur le marché du travail ne répartit pas la main-d'œuvre suivant la

confrontation entre l'offre et la demande comme le postule le modèle traditionnel, mais répartit les individus dans une structure déjà figée. Ces structures sont bien au cœur des résultats fournis par le modèle Log-linéaire.

Au-delà, ce qui reste fondamental dans le fonctionnement du marché du travail, c'est l'interdépendance de ses variables. Ce qui a conduit certains auteurs à préférer à la notion du marché du travail, celle de système, voulant traduire ainsi les multiples influences qui s'exercent sur le marché.

On a pu observer une féminisation du marché du travail qui va de pair avec le développement du tertiaire. Ce phénomène ne traduit pas une véritable mixité des emplois, puisque le marché du travail est sexuellement discriminé : les femmes occupent les emplois dans les métiers du tertiaire où elles sont majoritairement représentées tandis que les hommes sont surtout présents dans les métiers du secondaire. Cette discrimination est relativement stable dans le temps. Au demeurant, deux facteurs déterminants expliquent la féminisation des emplois : la nature des emplois (secondaire ou tertiaire) et le niveau de qualification. Dans ce dernier cas, on constate que le niveau de qualification requis pour l'emploi des femmes est en général faible et les emplois qualifiés qu'elles occupent sont spécifiques à certains types de services ou à certaines branches telles que le textile, l'habillement, etc.

Au total, les structures professionnelles jouent un rôle dans la féminisation des emplois : ce rôle est relativement plus favorable dans le tertiaire que dans le secondaire. Il y a donc en filigrane un "effet sectoriel" que nous n'avons pas cerné, la variable secteur d'activité n'étant pas prise en compte dans les données. En effet, *la féminisation de la plupart des métiers est plus importante dans les secteurs d'activité les plus féminisés.*

Outre le fait que les femmes excellent dans les métiers du tertiaire (métiers de la santé, métiers des soins personnels, personnel qualifié des soins para-médicaux et sociaux, etc.), il ne nous est pas permis de projeter de façon mécaniste ce schéma global à l'échelle régionale, qui reste notre unité spatiale. Les régions telles que la Franche-Comté, où ces métiers sont relativement peu représentés, apparaissent comme relativement les plus féminisées. A l'inverse, la Provence Alpes Côte d'Azur, bien que dotée majoritairement des métiers du tertiaire, se caractérise par une sous-féminisation relative. Ce paradoxe semble montrer que les structures professionnelles jouent un rôle moindre au niveau régional, du moins, par à

l'offre d'emploi des femmes. Cette appréciation sera plus judicieuse, grâce à une analyse de la composition sectorielle de l'offre d'emploi dans chaque région.

Il faut également voir dans ce paradoxe la traduction d'un *effet régional pur*, c'est-à-dire les caractéristiques particulières à chaque région : l'histoire et les structures économiques de chaque région induisent des spécificités se manifestant par l'attraction ou la répulsion vers tel ou tel type d'emploi. Dans les régions Nord-Pas-de-Calais et Lorraine par exemple, nous avons observé qu'il y avait une forte proportion de jeunes dans les métiers peu qualifiés du secondaire. On ne peut nier que ce phénomène soit lié à la vocation industrielle des deux régions. Par ailleurs, leur accession aux emplois peu qualifiés serait à rapprocher de leur niveau de formation relativement faible ; ce niveau de formation est influencé par leur origine sociale (la plupart des jeunes sont fils d'ouvriers).

D'autre part, les régions telles que la Provence alpes côte d'Azur ou le Languedoc-Roussillon, connaissent une sous-féminisation relative, qui traduit entre autres, les adaptations spécifiques qui se portent sur le fonctionnement du marché du travail, où les tensions sont assez vives.

Les particularités régionales vont donc conduire à un processus d'allocation du travail qui se réalise de façon non identique même pour des régions ayant une même structure d'emploi. C'est le cas notamment de l'Île de France et de la Provence alpes côte d'Azur, si l'on s'en tient à l'allocation de l'emploi tertiaire.

En définitive, les spécificités régionales témoignent de la nécessité de ne pas gommer la dimension spatiale dans le fonctionnement de marché du travail, ce que tend à occulter toute hypothèse d'homogénéité des comportements dans l'espace. Ce résultat valide l'une des critiques portées au modèle traditionnel du marché du travail. Cependant, on ne saurait associer le marché du travail à un espace limité géographiquement, puisque les marchés sont, historiquement générés, en relation les uns avec les autres. Les différenciations spatiales relèvent d'une logique « qui intègre des systèmes différents « d'usage » de la force de travail en fonction, notamment des conditions de l'environnement local ». Les modes de fonctionnement des marchés du travail sont donc spatialement différents.

S'il est clair que l'espace doit être pris en compte dans le fonctionnement du marché du travail par la médiation des structures, il reste que le cadre territorial choisi pour analyser l'allocation du travail peut influencer le résultat : « si le cadre choisi est trop vaste, l'observation utile peut être noyée au milieu d'une masse d'informations secondaires [...] si le cadre de référence est trop petit, on n'englobera pas la totalité des données, et l'analyse sera incomplète ou même fautive, exactement comme lorsqu'on sort une phrase de son contexte ».

Outre le fait que la région est une unité administrative et s'apparente au concept d'espace limité dans la problématique néo-classique, elle demeure un cadre pertinent pour l'analyse des échanges en travail. Il ne s'agissait pas dans ce travail de supposer que les régions sont des entités autonomes où les comportements sont homogènes, mais de les mettre en relation les unes avec les autres. Les effets régionaux mettent bien en évidence les différences existant entre les marchés locaux du travail, ainsi que leurs interrelations. L'outil d'analyse utilisé s'avère donc précieux.

Avant de clore ce travail, nous allons faire deux remarques terminales :

- outre les aspects structurels établis, l'instrument que nous avons utilisé ici pour l'étude de l'offre d'emploi s'impose dans des analyses qui privilégient une démarche exploratoire sur une démarche explicative. En somme, il pourrait s'agir d'une analyse se situant en amont de toute analyse de nature explicative (on constatera que dans les modèles économétriques traditionnels, le choix des variables à modéliser obéit le plus souvent à l'apriorisme et la place accordée à l'exploration des données est trop limitée), et qui permettrait ainsi de cerner les variables qui interviennent de manière significative dans l'explication d'un phénomène. Toutefois, un certain nombre d'auteurs tendent aujourd'hui d'approcher économétriquement les modèles Log-linéaires (GOURIEROUX, 1984 ; VERGER, 1983 ; LOLLIVIER, 1983). Dans cette optique, on explique par exemple la structure de l'emploi à travers l'effet d'un certain nombre de variables endogènes. Le développement de cette approche ne manquera pas d'intérêt et certains économètres y voient le moyen de remédier aux carences des modèles macro-économiques, du fait de son champ d'investigation à la fois macro et micro-économique.

- l'analyse que nous avons faite ne porte que sur les stocks. Or, une meilleure compréhension du marché du travail exige également l'analyse des flux. J.P. FITOUSSI souligne d'ailleurs

que « les flux d'embauche, de licenciement, par exemple, sont généralement quantitativement bien supérieurs aux stocks de chômeurs et de postes vacants ». L'analyse en terme de flux serait encore plus féconde grâce aux prolongements du modèle Log-linéaire, notamment sa formulation dans un tableau de contingence où il existe des variables ordinales.

Ces prolongements débouchent sur des modèles d'association qui sont des instruments particulièrement adaptés à l'analyse des mobilités sur le marché du travail. Ce point n'a pas bénéficié ici d'une application pratique pour des raisons qui tiennent surtout des contraintes techniques, notamment informatiques. En effet, les possibilités des logiciels que nous disposons actuellement ne nous ont pas permis d'envisager avec efficacité l'exploration des données par de tels outils. Tout de même, les investigations en cours de maturation depuis un certain temps leur ouvrent des perspectives d'utilisation fort prometteuses.

ANNEXES

NOMBRE D'ITERATIONS : 30 SEUIL = 0.0000

*** ATTENTION : TOUS LES RESULTATS SONT MULTIPLIES PAR 10**3 ***

ITERATION N	1	MARGE :	235	VMAX	=	1046.2583160814
ITERATION N	2	MARGE :	234	VMAX	=	102.5752366873
ITERATION N	3	MARGE :	345	VMAX	=	4.5029891088
ITERATION N	4	MARGE :	245	VMAX	=	0.9929700578
ITERATION N	5	MARGE :	234	VMAX	=	0.4332025987
ITERATION N	6	MARGE :	235	VMAX	=	0.1311085466
ITERATION N	7	MARGE :	245	VMAX	=	0.1178596123
ITERATION N	8	MARGE :	345	VMAX	=	0.0561051183
ITERATION N	9	MARGE :	234	VMAX	=	0.0476673433
ITERATION N	10	MARGE :	245	VMAX	=	0.0211026954
ITERATION N	11	MARGE :	235	VMAX	=	0.0119222677
ITERATION N	12	MARGE :	234	VMAX	=	0.0075696210
ITERATION N	13	MARGE :	245	VMAX	=	0.0045504965
ITERATION N	14	MARGE :	345	VMAX	=	0.0029033218
ITERATION N	15	MARGE :	234	VMAX	=	0.0017514590
ITERATION N	16	MARGE :	245	VMAX	=	0.0009551345
ITERATION N	17	MARGE :	235	VMAX	=	0.0008441808
ITERATION N	18	MARGE :	234	VMAX	=	0.0003917987
ITERATION N	19	MARGE :	245	VMAX	=	0.0002415148
ITERATION N	20	MARGE :	345	VMAX	=	0.0001906786
ITERATION N	21	MARGE :	234	VMAX	=	0.0000981514
ITERATION N	22	MARGE :	245	VMAX	=	0.0000570427
ITERATION N	23	MARGE :	235	VMAX	=	0.0000580320
ITERATION N	24	MARGE :	234	VMAX	=	0.0000248814
ITERATION N	25	MARGE :	245	VMAX	=	0.0000150704
ITERATION N	26	MARGE :	345	VMAX	=	0.0000140032
ITERATION N	27	MARGE :	234	VMAX	=	0.0000063885
ITERATION N	28	MARGE :	245	VMAX	=	0.0000038554
ITERATION N	29	MARGE :	235	VMAX	=	0.0000042395
ITERATION N	30	MARGE :	234	VMAX	=	0.0000017450

TOTAL ITERATIONS : 1155.1681617325

SCALAIRES

EPSS (0.0000) = 9.07383338583
 Z(0.0000) = 24.3016528126
 LN Z(0.0000)/W(0.0000) = 1000.0000000000
 7858.3991606585

EPS2

1563. -518. -420. -345. -125. -775. -616. 364. -199. -413. -759. 63. -209. -634. -164. -247.
 -1439. 796. -743. -523. 391. 289. 404. 1078. 614. 496. 350. 797. 608. 397. 636. 505.
 3575. 446. 492. 530. 661. 345. 404. 1078. 614. 496. 350. 797. 608. 397. 636. 505.
 178. 1659. 356. 444. 1107. 1335. 26. 68. 39. 31. 23. 51. 39. 25. 40. 37.
 226. 28. 31. 34. 42. 22. 26. 68. 39. 31. 23. 51. 39. 25. 40. 37.
 11. 105. 23. 28. 70. 1000. 26. 68. 39. 31. 23. 51. 39. 25. 40. 37.

EPS3

-1308. -417. -455. -1668. -1061. -1669. -1193. 145. -1935. 578. -153. -572. -1720. -267. 111. -1339.
 1553. -238. -305. 229. -1666. 366. 1515. 506. 87. 1074. 517. 340. 108. 461. 674. 158.
 163. 397. 382. 114. 209. 114. 183. 697. 87. 1074. 517. 340. 108. 461. 674. 158.
 2848. 475. 444. 758. 114. 869. 2741. 1659. 6. 78. 37. 24. 8. 33. 48. 11.
 12. 29. 28. 9. 15. 8. 13. 50. 6. 78. 37. 24. 8. 33. 48. 11.
 203. 34. 32. 54. 8. 62. 197. 1000. 6. 78. 37. 24. 8. 33. 48. 11.

EPS4

-73. 295. 195. 0. -183. -384. 25.
 906. 1310. 1185. 975. 812. 664. 1025.
 154. 224. 202. 167. 139. 114. 1000.

EPSS

243. -320. 38.
 1227. 699. 1039.
 636. 364. 1000.

bon nombre de jeunes doivent passer avant de retrouver un emploi stable, ce que JOSE ROSE (1984) appelle «transition professionnelle» c'est-à-dire une succession des moments de formation, d'emploi et de chômage. Ainsi les jeunes auraient plus de problèmes de stabilisation que d'insertion puisque leur "employabilité" est manifeste dans la plupart des métiers - surtout les métiers qui ont leur faveur- mais « s'ils (les jeunes) retrouvent plus vite que les autres un emploi, en revanche, l'emploi qu'ils occupent est plus souvent à durée limitée; leur insertion dans la vie professionnelle se fait souvent à travers des emplois instables; par ailleurs l'emploi précaire tendrait à appeler l'emploi précaire » [24]. Cette précarisation de l'emploi des jeunes, voire son instabilité reflète en quelque sorte le rôle dévolu à la main-d'oeuvre juvénile et qui est celui d'assurer la régulation du fonctionnement de l'économie : « la régulation s'assimile ici à la nécessité d'obtenir à chaque instant une main-d'oeuvre adaptée en volume et en qualité aux besoins de l'appareil de production et à réserver » à la marge un certain volume de main-d'oeuvre prête à être immobilisée dans les emplois ». On voit implicitement apparaître le rôle de main-d'oeuvre de réserve jadis joué par les femmes à l'occasion des transformations du système productif opérées depuis la décennie soixante.

La fonction de régulation dévolue aux jeunes est soucieuse des recommandations du IV^e plan : adapter la structure qualitative de la main-d'oeuvre aux besoins de l'économie. Ainsi, l'offre d'emploi des jeunes dépend de l'état de la conjoncture et on peut se demander si dans l'état actuel de crise qui implique une raréfaction de l'emploi les mesures prises dans le cadre des pactes pour l'emploi par exemple ont une efficacité pour leur insertion. De toutes les façons le taux de demande d'emploi des moins de 25 ans n'a cessé d'augmenter pour atteindre 40% en 1981.

D'autre part, les stages en entreprises ou les contrats d'emploi-formation institués par les pactes pour l'emploi apparaissent comme de nouvelles formes d'emploi précaire dans la mesure où les jeunes ont également de difficultés d'insertion à l'issue de leur stage ou contrat. Ainsi que le souligne G. MALGLAIVE [25] « proclamé comme un moyen de donner aux jeunes "un métier pour réussir", le dispositif mis en place n'en conduit que très peu à l'emploi ».

Une enquête effectuée par le CEREQ [26] auprès des jeunes stagiaires de la première campagne indique qu'un jeune sur cinq est encore en stage 16-18 ans un an et demi après être entré dans le dispositif, un cinquième des jeunes a un emploi dont le quart est en

apprentissage, 45% sont au chômage et 8,5% en inactivité. On constate donc que les résultats sont assez minces bien que ces dispositifs leur permettent une réelle expérience professionnelle où ils affermissent leur orientation et contribuent progressivement - même de façon limitée - à les intégrer sur le marché de travail. Sur le plan politique, ces mesures ont permis de contenir le chômage, vu le nombre de jeunes inscrits dans les stages de formation alternée ou retournés ou maintenus en formation initiale (l'effectif des bénéficiaires d'un stage rémunéré est passé de 140.000 en 1982 à 248.000 en 1985 et parmi ces stagiaires 54.000 sont considérés comme ayant une activité professionnelle et correspondant aux T.U.C.). D'autre part, en terme de chômage évité par ces mesures, ce sont 274.000 jeunes qui ne sont pas en chômage en 1983 contre 148.000 en 1981.

La limite du dispositif tient également du fait qu'il se heurte aux contraintes d'adaptabilité des entreprises en période de crise, et dont les jeunes sont surtout victimes. En effet, afin de restaurer les mécanismes concurrentiels et par conséquent la compétitivité pour favoriser la demande, les entreprises pratiquent la politique de flexibilisation de la main-d'oeuvre pour adapter celle-ci au rythme de la production. Cette politique est aussi plus facilement atteinte que la multiplication des statuts intermédiaires s'y prêtent et les avantages qu'en tirent les entreprises ne sont pas minces : la différenciation des statuts entraîne des rémunérations moins élevées dans la mesure où l'on ne prend pas en compte l'ancienneté ; elle réduit les avantages sociaux et rend les conditions de travail moins favorables. De surcroît, la politique de flexibilisation de la main-d'oeuvre amoindrit les coûts salariaux et développe pour les entreprises les aptitudes à changer de postes de travail [27]. On voit alors que les mesures prises pour l'insertion des jeunes se situent entre deux objectifs contradictoires : contenir d'une part le chômage et faciliter l'adaptation du système productif d'autre part, ce qui amène M. AGLIETTA à conclure que « la politique de l'Etat viserait alors contradictoirement à intensifier la mobilité forcée et à limiter les conséquences sociales du développement du chômage et l'insécurité. Et ceci expliquerait deux des caractéristiques actuelles des politiques d'emploi : le fait qu'elles sont essentiellement des politiques qui organisent les conditions générales de la sélection pour le compte des employeurs et le fait qu'elles sont fragmentaires » [28].

Cette analyse doit être complétée par une autre qui permette de saisir comment l'âge permet de structurer les régions.

7.3.3 L'âge comme facteur de structuration des salariés suivant les régions

L'objet du présent paragraphe, c'est de prospecter la répartition spatiale des salariés par classe d'âge et d'essayer de dégager les caractéristiques comportementales des différentes catégories d'une région à une autre. L'analyse que nous conduirons permet d'apprécier la différenciation spatiale des salariés à travers leur âge et qui serait induite non seulement par les facteurs socio-économiques, mais également démographiques, lesquels reflètent en quelque sorte ou en partie la spécificité voire la dynamique des marchés de travail régionaux. C'est ainsi que les régions aux actifs "jeunes" sont celles du Nord où la fécondité, relativement élevée se superpose à la situation dégradée de l'emploi y renforçant les difficultés d'accès à l'emploi alors que celles du Sud, relativement "âgées" se caractérisent par une fécondité faible, une situation d'emploi relativement bonne ; la conjonction des deux facteurs permettant d'atténuer les chocs possibles sur leurs marchés de travail. Toutefois la part des différentes catégories ne semble pas avoir sensiblement modifié d'une région à l'autre malgré les constats qui viennent d'être faites, ce qui prouve que d'autres facteurs interviennent sur le plan régional pour dynamiser les marchés de travail. Nous verrons apparaître le clivage Nord-Sud dans la répartition des salariés, ce clivage étant tempéré par des facteurs que nous explorerons ensuite avant de dresser une typologie des régions.

7.3.3.1 Une opposition Nord-Sud : le Nord "jeune" et le Sud "vieux"

Une vision schématique de l'effet de structure région-Age permet d'établir un clivage des régions françaises selon les catégories d'âge de leurs salariés. Aussi peut-on distinguer deux grands ensembles de régions :

- la partie Nord, formée des régions allant des pays de la Loire et en passant par le Nord-Pas-de Calais jusqu'en Franche Comté. On y trouve une forte proportion des jeunes salariés même si la représentativité de ceux-ci est légèrement différenciée entre ces régions : alors que les deux premières ont une part relativement supérieure à la moyenne des moins de 25 ans et des salariés de 25-31 ans, la Franche Comté se situe quant à elle à un niveau relativement moyen. Peut-être est-ce dû à un effet d'échantillonnage dans la mesure où le poids des effectifs salariés de cette région est faible dans l'échantillon ?

On remarquera que ce groupe de régions correspond à ce que les géographes ont en d'autres circonstances dénommé le "croissant fertile", l'expression reflétant les taux de fécondité assez élevés que l'on y enregistre [29] : Nord-Pas-de Calais (32,8%), pays de la Loire (31,9%), Franche Comté (31%). Ainsi, on peut dans un premier temps estimer que la forte représentation des jeunes dans ces régions est expliquée par les facteurs démographiques, bien que ceux-ci expriment un effet de structure à associer plus à la demande d'emploi qu'à l'offre d'emploi.

- la partie Sud enregistre plutôt une forte représentativité des salariés adultes et une sous-représentativité des salariés jeunes. On peut aussi lier partiellement ce phénomène à la faible fécondité des régions qui composent cette partie : Provence-Alpes-Côte d'Azur (16,5%), Auvergne (16,7%), Languedoc-Roussillon (17,5%), Limousin (20,4%). Cette faible fécondité entraîne donc un renouvellement assez limité de la main-d'oeuvre juvénile et il n'est pas étonnant de retrouver une structure d'emploi plus favorable aux salariés des classes d'âge élevées qui y sont fortement représentés, notamment les salariés des classes d'âge 48-54 ans et 55 ans et plus. Toutefois, la fécondité n'expliquerait pas elle seule cette représentativité, le phénomène de "retour au pays" vers l'âge de la retraite voire à cet âge l'amplifiant, phénomène qui a été favorisé par le cadre de la vie et l'attractivité de ces régions : « il reste que la recherche d'un meilleur cadre de vie explique une part significative de la mobilité comme le montrent l'évolution récente de l'urbanisation, l'accentuation de l'attractivité de la quasi-totalité des départements situés au Sud de la Seine et le développement des migrations de retour au pays» [30].

Cette opposition Nord-sud reste valable du point de vue de la formation acquise par les salariés, laquelle peut structurellement expliquer la forte ou la faible représentativité de ceux-ci dans une région quelconque.

Les régions du Nord (à l'exception de l'Ile-de-France) ont une faible proportion de jeunes pouvant accéder à un niveau de formation élevé ce qui n'est pas le cas des régions du Sud. En effet, beaucoup parmi les jeunes des régions Nord-Pas-de Calais, Picardie, Champagne-Ardenne, Haute-Normandie, Basse-Normandie ont un niveau de diplôme ne dépassant pas le B.E.P.C. (plus de 48% pour une moyenne nationale de 44%) [31]. A l'inverse, les régions du Sud - surtout celles où siègent les grandes métropoles - sont caractérisées par un fort pourcentage de bacheliers et de jeunes ayant un diplôme d'étude supérieure. Ce résultat découlerait de la structure sociale qui est spatialement typée. Les

régions du Nord sont en général industrialisées avec une forte proposition d'ouvriers. Or, un certain nombre d'études ont montré que les jeunes issus des milieux ouvriers sont moins diplômés en moyenne que ceux des milieux des cadres, d'employés etc.

Du fait de leur formation relativement peu élevée et par conséquent leur faible niveau de qualification, les jeunes des régions du Nord entrent très tôt sur le marché du travail, où la structure des emplois leur est d'ailleurs assez favorable -ce sont des emplois industriels n'exigeant pas un niveau de qualification élevé- L'existence d'emplois peu qualifiés - des métiers peu qualifiés- fortement supérieurs à la moyenne nationale dans ces régions associée au faible niveau de formation semble montrer que la perception du marché local influe sur l'entrée plus ou moins tardive des jeunes sur le marché. Autrement dit, au vu de la structure des emplois offerts et l'acuité des tensions manifestes sur les marchés du travail de ces régions, les jeunes auraient tendance à sortir vite du système scolaire pour entrer dans la vie active, ce d'autant plus que les régions en question, qui ont été durablement affectées par la crise, présentent des perspectives d'avenir incertaines. Ces comportements, outre les facteurs démographiques, sont donc à même de justifier une part très importante des effectifs salariés de moins 25 ans dans les régions de la moitié Nord.

A l'inverse dans le sud, il existe une faible proportion d'emplois (de métiers) industriels exigeant un niveau de qualification peu élevé. Ce sont plutôt les emplois (métiers) tertiaires qui y priment ; ils exigent de plus un haut niveau de qualification. On comprend alors qu'il faudra plus de temps aux jeunes pour acquérir une formation répondant aux exigences des emplois offerts ; leur entrée tardive sur le marché peut expliquer leur sous-représentativité dans les régions de la moitié sud.

Donc la structure des emplois offerts joue au niveau régional un rôle fondamental dans l'accession à l'emploi des jeunes, et tout se passe comme s'ils adaptaient leur formation et par conséquent la durée de celle-ci, aux types d'emplois présents dans leur région, en vue d'une meilleure insertion professionnelle.

Sur la base des analyses qui viennent d'être faites, on relève que la proportion des jeunes salariés est forte dans les régions où il existe une part importante des métiers du secondaire (d'emplois du secondaire). C'est le cas notamment des régions telles que : Nord-Pas-de Calais, Lorraine, Picardie, Champagne-Ardenne etc.

Tableau 26 : Discrimination région-âge selon le sexe

| | Ages→ | - 25 | 25-31 | 32-39 | 40-47 | 48-54 | 55 + |
|----------------------|-------|------|-------|-------|-------|-------|------|
| Régions ↓ | | | | | | | |
| Ile de France | | + | 0 | | 0 | | |
| Champagne | | 0 | 0 | | 0 | | -- |
| Picardie | | 0 | 0 | + | 0 | - | -- |
| Haute Normandie | | | 0 | ++ | 0 | -- | -- |
| Centre | | -- | + | 0 | 0 | | - |
| Basse Normandie | | - | | - | ++ | - | ++ |
| Bourgogne | | -- | ++ | + | -- | | |
| Nord-Pas-de-Calais | | - | 0 | - | 0 | 0 | 0 |
| Lorraine | | 0 | 0 | - | 0 | + | |
| Alsace | | -- | | | ++ | | ++ |
| Franche comté | | | | ++ | | | -- |
| Pays de la Loire | | - | 0 | + | + | | |
| Bretagne | | | | + | | | + |
| Poitou Charentes | | | | | - | - | |
| Aquitaine | | - | - | + | + | 0 | ++ |
| Midi-Pyrénées | | - | | - | - | ++ | + |
| Limousin | | | -- | + | -- | ++ | |
| Rhône-Alpes | | ++ | 0 | 0 | 0 | 0 | 0 |
| Auvergne | | 0 | - | 0 | + | | ++ |
| Languedoc-Roussillon | | | 0 | | + | - | - |
| Prov.-Alpes-C.A.-CO | | - | 0 | - | + | 0 | |

++ : croissance très forte

+ : croissance forte

0 : stabilité

- : décroissance

-- : décroissance très forte

A l'inverse, cette proportion reste faible dans les régions où la part des métiers du tertiaire est assez élevée. Ce résultat ne confirme pas totalement les tendances qui ont été perçues globalement puisque les jeunes ont également la faveur de certains métiers du tertiaire. Cela peut être lié au fait que la région constitue un cadre d'analyse assez vaste et qu'elle masque certains phénomènes qui seraient davantage plus perceptibles à un niveau d'analyse plus fin, notamment au niveau d'un département ou d'une commune. Toutefois on ne peut ignorer que les résultats obtenus traduisent l'existence d'un effet de structure dû à la composition de l'offre d'emploi régional ; cet effet explique en partie les différences de comportement des salariés des différentes catégories d'âge sur les marchés du travail régionaux.

On relèvera la similitude des comportements des jeunes avec ceux des femmes au niveau régional puisque la forte participation des jeunes à l'activité va de pair avec l'industrialisation de la région, ainsi que nous l'avons observé chez les femmes.

Cependant la dichotomie qui vient d'être esquissée et qui est en parfaite adéquation , d'une part sur le plan démographique à la division séculaire, d'autre part à la division du point de vue de la formation, doit être relativisée, car les mouvements sociaux des années soixante et des années soixante-dix l'ont entamée avant que la crise ne vienne continuer à la façonner. Nous serons donc amenés à essayer de reconstituer le nouveau paysage régional des différentes catégories d'âge en invoquant d'autres facteurs qui les déterminent, bien que les tendances générales ne disparaissent pas pour autant : le Nord reste toujours relativement "jeune" et s'oppose au Sud qui , tout en étant "vieux", a tendance à rajeunir.

7.3.3.2 Autres déterminants de la fracturation spatiale des salariés suivant les classes d'âge

Il vient d'être établi au paragraphe précédent que l'offre d'emploi est segmentée au niveau régional et que cette segmentation peut être liée à la composante structurelle de l'emploi dans chaque région. Bien entendu, cette composante structurelle est favorisée ou pas par les facteurs comme la démographie, la formation etc. Nous nous sommes situés à ce stade dans une vision statique des différents marchés. Une analyse dynamique de ceux-ci, du moins sur la période étudiée montre l'intervention d'autres facteurs tels que les mobilités spatiales, dont le rôle, quoique semble-t-il relativement limité, dans la régulation des différents marchés du travail régionaux, devrait être mentionné.

L'examen du tableau 26 indique une quasi-stabilité des effectifs salariés jeunes dans la région Nord-Pas-de Calais et à un degré moindre en Lorraine. Cette relative stabilité contraste avec le processus de grande ampleur amorcé dans ces régions où la crise frappe durablement les secteurs industriels traditionnels, les industries extractives, la sidérurgie et le textile qui y constituent des activités dominantes. Par conséquent, la diminution de l'emploi, conjuguée à la pression démographique aurait davantage accentué les déséquilibres du marché, aggravant par la même occasion la situation des jeunes sans expérience qui y sont largement représentés et qui pour la plupart entrent sur le marché du travail pour la première fois. Certes, peut-être pourra-t-on justifier cette relative stabilité par des entrées massives en chômage, puisque la région se situe parmi celles où le taux de chômage est très élevé (près de 14% en 1985). Mais il y a un autre facteur qui joue : les migrations interrégionales. Celle-ci sont liées à l'âge, les jeunes (les moins de 25 ans) ayant la faculté à émigrer plus que les autres tranches d'âge. En effet sur les marchés du travail des régions Nord-Pas-de Calais par exemple, les jeunes en concurrence avec les actifs plus âgés et plus qualifiés, émigrent vers les métropoles des régions fortement urbanisées et attractives telles que l'Ile-de-France, la Provence-Alpes-côte d'Azur-Corse, le Rhône-Alpes, où le développement du tertiaire offre, entre autres, une structure d'emploi qui leur est favorable (emplois tertiaires à fort taux de renouvellement de la main-d'oeuvre , cf. 7.3.1) [32]. Ce résultat est d'autant plus significatif que l'on note une évolution positive des effectifs des jeunes en Rhône-Alpes et en Ile-de-France (cf. tableau 26), et ce sur la période 1976-1981. Il est à signaler que le taux d'émigration qui est de 6,2% au recensement de 1982 en Nord-Pas-de Calais et qui touche surtout les jeunes, donne l'idée sur la propension à quitter la région. Ainsi, pour le jeune, la migration peut constituer une condition de son insertion professionnelle.

On peut faire un parallélisme entre l'évolution des effectifs des moins 25 ans au Centre et en Bourgogne et l'orientation des migrations spatiales, du moins pour ce qui concerne les jeunes surtout, car elles ne sont pas neutres des caractéristiques du marché. En effet, malgré une forte croissance de l'emploi global au Centre sur la période 1975-1982 [33], on y enregistre une très forte baisse des emplois des jeunes ; l'ampleur de cette baisse est à noter également en Bourgogne (cf. tableau 26). On peut alors imaginer que ces baisses s'opèrent au profit des régions Ile-de-France, Rhône-Alpes, du fait de la connexité spatiale du Centre avec l'Ile-de-France d'une part , de la Bourgogne avec le Rhône-Alpes et l'Ile-de-France d'autre part. Il faudra néanmoins relativiser ce processus, car même s'il reste vérifié, l'évolution de

l'emploi des jeunes dans les deux régions que sont l'Ile-de-France et le Rhône-Alpes, ne saurait être imputable au seul facteur migratoire des régions Centre et Bourgogne.

L'évolution des salariés de moins 25 ans en Alsace apparaît fort surprenante (cf. graphique 25). D'une représentation très supérieure à la moyenne en 1976, cette catégorie a vu ses effectifs baisser régulièrement sur la période pour se rapprocher de la moyenne. Et pourtant, la région présente une texture économique favorable au développement d'emplois adaptés aux jeunes. De plus, le taux de chômage de l'Alsace se situe en dessous de la moyenne nationale en 1984 (8,5% pour une moyenne nationale de 10,5%) et le taux de chômage des jeunes dans la région se situe à un niveau moyen. On ne pourrait pas tellement imputer ce phénomène à la migration. En effet, l'Alsace étant une région presque fermée sur elle-même (on ne quitte pas volontiers la région qui occupe parmi les régions Françaises la dernière position en matière d'émigration ; elle est aussi en avant dernière position en matière d'immigration) [34], on peut lier cette baisse des effectifs à l'insuffisance du renouvellement des ressources humaines puisqu'elle connaît une baisse de la fécondité depuis la décennie soixante-dix. Dans ce cas, l'évolution de l'emploi global qui y a été enregistrée profite surtout aux tranches d'âge 40-47 ans, 55 ans et plus où la croissance des effectifs est très forte, entre 1976 et 1981 ; les autres tranches d'âges 25-31 ans, 32- 39 ans) ont plutôt des effectifs relativement stables.

Les régions dotées de très grandes métropoles sont caractérisées par une sur-représentation des tranches d'âge intermédiaires (32-39 ans, 40-47 ans, 48-45 ans). C'est le cas évidemment de l'Ile-de-France, du Rhône-Alpes et de la Provence-Alpes-Côte d'Azur Corse. Dans les deux premières, on peut y voir l'effet d'une immigration forte des effectifs - cet effet est aussi valable pour la troisième région- qui viennent y travailler jusque l'âge adulte (ce qui explique leur forte proportion) avant d'amorcer le mouvement de retour vers les régions d'origine à proximité de l'âge de la retraite. La part relativement importante des salariés de 55 ans et plus en Ile-de-France serait à prendre avec prudence, à moins que cela ne reflète la sortie très tardive des salariés des métiers de très haut niveau de qualification, dont la plupart appartiennent à cette tranche d'âge et dont on a aussi relevé leur polarisation surtout en Ile-de-France.

Quant aux régions telles que la Bretagne, la Basse-Normandie, le Poitou Charentes, qui ont déjà une forte proportion de jeunes salariés, elles voient également leur part de 55 ans et plus se situer nettement au-dessus de la moyenne. Ces régions sont parmi celles qui ont

bénéficié de la décentralisation industrielle. L'effet de celle-ci est évidemment l'atténuation de l'exode rurale qui s'était longtemps effectuée au profit de l'Ile-de-France. De plus, elle peut avoir influencé sur l'inversion des flux migratoires de façon à ce que l'on interprète la forte représentativité des 55 ans et plus comme un phénomène de retour dans les régions.

Ces analyses nous sortent donc du schéma global où nous avait plongé la dichotomie Nord-sud, en ce qui concerne l'emploi salarié régional ventilé suivant les classes d'âge. On peut alors s'en inspirer pour proposer une typologie plus fine des régions par âge.

7.3-3.3 Typologie régionale des salariés selon les tranches d'âge

Il est possible de distinguer plusieurs groupes de régions à partir des observations découlant des analyses que nous venons de mener précédemment (cf. carte 4) :

- les régions à sur-représentativité des effectifs salariés aux âges intermédiaires (25-54 ans) et à leur sous-représentativité aux âges extrêmes (moins de 25 ans, plus de 55 ans) : Rhône-Alpes et Ile-de-France, bien que cette dernière ait une forte proportion des plus de 55 ans. Il s'agit, bien entendu de deux régions où coexistent une part relativement plus importante des emplois tertiaires et une part relativement plus importante des emplois industriels également, et ce grâce à l'équilibrage de leur texture économique. Ces emplois sont pour la plupart d'un haut niveau de qualification (cf. 7.1.2) et l'on peut par là comprendre qu'ils soient plutôt adaptés aux classes d'âge intermédiaires, lesquelles ont déjà acquis une certaine expérience professionnelle que les jeunes sont encore en manque. Cette structure d'emploi favorable aux salariés de classes d'âge intermédiaires explique en grande partie leur représentativité forte dans ces régions. Cependant celles-ci cumulent aussi des avantages pour les jeunes malgré la part importante que y représentent les emplois qualifiés, qui leur sont qui leur sont généralement fermés. En effet, elles sont dotées d'une texture économique assez équilibrée, qui leur offre des emplois peu qualifiés aussi bien dans les métiers du tertiaire (métiers non qualifiés du commerce et distribution, métiers non qualifiés de l'hôtellerie etc.) que les métiers du secondaire (ouvriers spécialisés et manoeuvres). Ces métiers sont ceux où existe un fort "turnover" (la forte activité du marché du travail ici, caractérisée par une forte mobilité se traduisant par des sorties aisées du chômage qui compensent de nombreuses entrées semble l'attester) et par lesquels les jeunes font leurs premières expériences professionnelles. C'est pour cela qu'elles constituent pour eux des points d'attraction au

moment de leur insertion initiale, lorsque les conditions d'accèsion à l'emploi sont souvent défavorables dans leur région d'origine ; il semble qu'après une période d'instabilité, leur affectation aux emplois stables leur permet d'y rester -peut-être jusqu'à l'âge proche de la sortie d'activité- avant d'amorcer le processus de retour aux régions d'origine, ce qui expliquerait la représentativité des tranches d'âge intermédiaires dans ces deux régions.

-les régions où coexistent la sur-représentativité des classes d'âge extrêmes et la sous-représentativité des classes d'âge intermédiaires : Basse-Normandie, Bretagne, Bourgogne, Poitou-Charentes. On constate que ce groupe de régions a un profil qui s'oppose à celui du groupe précédemment défini. Il s'agit des régions à taux de fécondité comme il a été déjà mentionné auparavant. D'autre part, la proportion d'emplois de faible niveau de qualification est importante puisque ce sont des régions, soit de tradition industrielle ancienne (Centre, Bourgogne), soit de décentralisation (Bretagne, Poitou-Charentes); la décentralisation a permis de créer sur place une part importante d'emplois qui requièrent des techniques banalisées. Cette conjonction de facteur explique en partie la sur-représentativité des salariés de moins de 25 ans, car la structure d'emplois leur est favorable.

-l'émigration qui affecte ces régions du fait de l'ampleur des déséquilibres sur leurs marchés du travail, se refléterait dans la sous-représentativité des classes d'âge intermédiaires ; le vieillissement aux classes d'âge élevées dont on observe une forte proportion ici est dû aux retours des salariés vers ces régions. La Bretagne sur ce point obtient la palme d'or.

-les régions où il y a une sous-représentativité des salariés jeunes et une forte représentativité des salariés des tranches d'âge élevées : Languedoc-Rousillon, Midi-Pyrenées, Provence-Alpes-Côte d'Azur-Corse, Auvergne auxquelles on peut inclure à un degré moindre le Limousin dont la classe d'âge 40-47 ans est sous-représentée. On distingue ainsi deux régions fortement rurales (Limousin, Auvergne), trois régions d'urbanisation moyenne (Midi-Pyrenées, Aquitaine, Languedoc-Rousillon) et une région fortement urbanisée (Provence-Alpes-Côte d'Azur-Corse).

Tableau 27 : Discrimination region-age selon le sexe

| Régions ↓ | Ages→ | - 25 | 25-31 | 32-39 | 40-47 | 48-54 | 55 + |
|----------------------|-------|------|-------|-------|-------|-------|------|
| Ile de France | | | | | + | 0 | - |
| Champagne | | | | | -- | | - |
| Picardie | | -- | + | -- | ++ | + | |
| Haute Normandie | | + | | | ++ | ++ | -- |
| Centre | | - | + | + | 0 | + | - |
| Basse Normandie | | - | - | + | + | + | -- |
| Bourgogne | | + | | - | - | + | - |
| Nord-Pas-de-Calais | | 0 | + | - | - | - | + |
| Lorraine | | - | - | 0 | - | | |
| Alsace | | | - | + | + | + | + |
| Franche comté | | + | + | - | - | - | + |
| Pays de la Loire | | + | - | - | + | + | |
| Bretagne | | | | | | | ++ |
| Poitou Charentes | | + | - | | - | | -- |
| Aquitaine | | + | 0 | | -- | -- | |
| Midi-Pyrénées | | -- | + | | - | | |
| Limousin | | | | - | + | | |
| Rhône-Alpes | | | - | + | - | - | |
| Auvergne | | | | + | | + | - |
| Languedoc-Roussillon | | | + | | | | - |
| Prov.-Alpes-C.A.-CO | | 0 | - | - | - | + | + |

++ : croissance très forte

+ : croissance forte

0 : stabilité

- : décroissance

-- : décroissance très forte

On constatera donc que malgré l'idendité des profils qu'elles ont, ces régions ne disposent pas les mêmes structures d'emploi. Ce sont sûrement les facteurs démographiques conjugués à l'existence d'une proportion relativement importante d'emplois agricoles qui expliqueraient la sous-représentativité des salariés jeunes dans le Limousin et en Auvergne.

La Provence-Alpes-Côte d'Azur-Corse reste très favorable aux jeunes pour les mêmes raisons évoquées en ce qui concerne le Rhône-Alpes et l'Île-de-France. Cependant elle a une particularité par rapport aux deux régions ci-indiquées, ce qui reste prévisible dans la superposition de ses deux principales caractéristiques : celle d'être une région à fort degré d'urbanisation, canal par lequel se développent de nombreux emplois tertiaires et celle d'être aussi une région "vieille", du fait non seulement des facteurs démographiques, mais aussi ceux liés aux migrations interrégionales des actifs. Ces éléments ont des liens avec la structure d'emplois offerts dans la région et c'est pour cette raison que non seulement les salariés de plus de 55 ans y sont largement représentés mais aussi ceux des classes d'âge intermédiaires. On note d'ailleurs une représentativité des différentes classes d'âge, positivement liée à l'âge.

-les régions à forte représentativité des jeunes salariés et à sous-représentativité des autres catégories de salariés : Picardie, Champagne-Ardenne, Haute-Normandie, Nord-Pas-de-Calais, Alsace, Lorraine, Franche comté, Pays de la Loire. Il s'agit encore là des régions jeunes et de surcroît industrialisées. Elles ont de ce fait développé des emplois aux caractéristiques de la main-d'oeuvre juvénile en majorité.

Le Nord-Pas-de-Calais se singularise par un profil où la représentativité des classes d'âge décroît avec l'âge (ce profil est totalement opposé à celui de Provence-Alpes-Côte d'Azur-Corse). Ceci traduit la très forte juvénilité de la région. Mais en faisant un parallélisme avec les phénomènes migratoires, peut-on conclure que le "retour au pays" est de faible ampleur ici ? Il reste que, malgré la crise qui frappe la région, la structure d'emplois est encore favorable en région Nord-Pas-de-Calais aux jeunes. On pourra y voir également l'effet de son urbanisation moyenne, qui permet de développer les activités de service en régulière expansion, lesquelles offrent un éventail d'emplois relativement large, notamment pour les jeunes.

Au-delà de cette typologie, se dégagent quelques observations corroborant les résultats que nous révèle la discrimination sexuelle relative par âge des régions (cf. tableau 27) : les régions comportant une forte proportion de jeunes sont également celles où la part des

femmes de cette tranche d'âge est largement représentée. Ce résultat est bien entendu l'illustration de la composante structurelle de l'offre d'emploi.

De plus, le parallélisme tracé entre les caractéristiques de l'emploi des jeunes et celles de l'emploi des femmes se recoupe à l'échelon régional : les femmes et les jeunes excellent dans les emplois de faible degré de qualification et dans les régions où ceux-ci sont sur-représentés, la féminisation de l'emploi reste le fait de l'existence plus ou moins prononcée d'une structure d'emploi qui profite surtout aux femmes de moins de 25 ans.

-dans les régions urbanisées, la part des femmes est moins importante chez les moins de 25 ans ; elle est à l'inverse moyenne ou significative dans les tranches d'âge intermédiaires. Ce résultat est à rapprocher de celui obtenu quand on fait la discrimination sexuelle relative professionnelle (cf. 7.2) où il ressort qu'il y a un processus de féminisation des emplois dans les métiers qualifiés tels que personnel qualifié des services para-médicaux et sociaux, professeurs, professions littéraires et de l'information, cadres moyens etc. Ces métiers sont surtout polarisés dans les grandes métropoles ; ils exigent des études généralement longues et par conséquent une entrée en activité tardive. C'est ce qui expliquerait que l'on y retrouve une part importante des femmes des tranches d'âge intermédiaires.

CONCLUSION

L'analyse que nous venons de faire ici met en lumière le rôle structurant de l'offre d'emploi salarié. En effet il est apparu que différentes catégories de salariés avaient des comportements variant d'une profession à une autre ou d'une région à l'autre etc. , il y a même une interaction de ces variables. Ainsi en ce qui concerne la féminisation des emplois, on constate qu'il y a une concentration des femmes dans les métiers du tertiaire de faible niveau de qualification. Cela n'implique pas pour autant que ce résultat soit transposable au niveau régional. En effet, la région est une réalité complexe et rarement homogène. On a pu y relever quelques professions dominantes. Mais la diversité des évolutions qu'on y observe montre que

que celles-ci ne sont pas liées à l'implantation des métiers ou celle de l'activité dominante. Peut-être est-ce le reflet des statistiques globales que l'on saisit au niveau régional ? L'affinement de ces statistiques à une échelle encore plus fine telle que la zone d'emploi par exemple indique que la Provence-Alpes-Côte d'Azur, région la plus tertiarisée comprend des

zones industrielles (Fos, Marignane etc.) offrant des emplois aux jeunes et fermées aux femmes ; elles s'opposent aux zones tertiaires (Alpes-Maritimes, Var) où les actifs sont âgés et où la structure d'emplois est plus favorable aux femmes. Il apparaît là un ensemble de résultats intéressants, liés à une variété de situations géographiques que l'entité régionale gomme sensiblement.

Au-delà des résultats qui découlent des analyses faites ici, il reste que dans l'ensemble, c'est le tertiaire qui offre encore des possibilités des créations d'emplois aussi bien au niveau des métiers que des régions sur la période 1976-1981, bien que ses tendances s'infléchissent par rapport à celles enregistrées avant la crise (le taux de croissance annuel moyen est de 1,6% entre 1973 et 1982 contre 2,2% entre 1969 et 1974). Ces inflexions reflètent les mutations qui ont touché l'ensemble de système productif au cours de la période et qui ont contribué à façonner considérablement les conditions et de développement des emplois tertiaires, plaçant ceux-ci dans un environnement peu favorable. Certes, même à un rythme moins soutenu, les métiers du tertiaire ont permis de limiter les effets de la crise et d'atténuer les déséquilibres existants sur le marché du travail où la population active ne cesse de croître dans des proportions supérieures à celles qu'offrent les possibilités de créations d'emplois. La population active devrait encore augmenter jusqu'à l'an 2000 où il atteindra 26,5 millions d'actifs contre 24,2 millions en 1985. Le rôle relatif du tertiaire dans l'atténuation de la crise serait alors remis en cause dans la mesure où son essoufflement ne saurait être dissocié du déclin de l'industrie et par conséquent, il ne pourrait vraisemblablement plus être à même d'endiguer les déséquilibres sur le marché du travail si les conditions d'une reprise de l'activité globale ne sont pas réunies. Même dans les métiers de la santé où le développement de l'emploi a été important, les évolutions futures subiront les contrecoups de la crise.

En effet, les créations d'emplois dans ces métiers sont liées aux dépenses de la santé. Elles devraient continuer à s'infléchir comme on l'a déjà observé à partir de 1979 où il y a tassement des effectifs. Le taux de croissance annuel moyen par habitant des dépenses budgétaires de la santé, passé de 6,7% entre 1973 et 1979 à 5,5% entre 1979 et 1985 devrait se situer au niveau de 3,6% entre 1985 et l'an 2000. Ce ralentissement est dû en grande partie à la conjoncture défavorable dont les conséquences sur les emplois tertiaires seront négatives. Au niveau régional, la continuation de la crise a profondément modifié les conditions socio-économiques dans chaque région et les tendances précédemment observées y sont également

vérifiées, bien que diversifiées. L'essentiel des emplois créés dans les régions sur la période 1976-1981 (ce sont surtout les emplois tertiaires reste le fait des emplois publics) cas par exemple des emplois des métiers de la santé lesquels sont une diffusion spatiale relativement plus homogène. Le rôle des pouvoirs publics dans le développement de ces emplois est manifeste puisque les "différents transferts ont accentué la redistribution spatiale des revenus et permis une diminution des disparités régionales". C'est l'un des éléments qui ont favorisé le développement des emplois tertiaires dans les régions rurales telles que le Limousin, l'Auvergne, etc., leur permettant de combler une partie de leur retard sur les autres régions.

Les régions qui ont bénéficié de la décentralisation ont mieux supporté les effets de la crise. Elles tirent également parti des coûts salariaux plus faibles qu'ailleurs qu'impliquent les emplois de faible niveau de qualification qui y ont été décentralisés.

Les effets de la conjoncture qui affectent le tertiaire s'ils se pérennisaient, aggraveraient davantage les difficultés des régions du Nord et Nord-Est (Nord-Pas-de Calais, Lorraine, Champagne-Ardenne etc.) où les problèmes de reconversion ont déjà engendré de nombreuses suppressions d'emplois.

Somme toute, l'autonomie du tertiaire par rapport à la conjoncture doit être relativisée et l'on ne pourra pas longtemps tabler sur les métiers de ce secteur pour préserver l'équilibre sur le marché du travail.

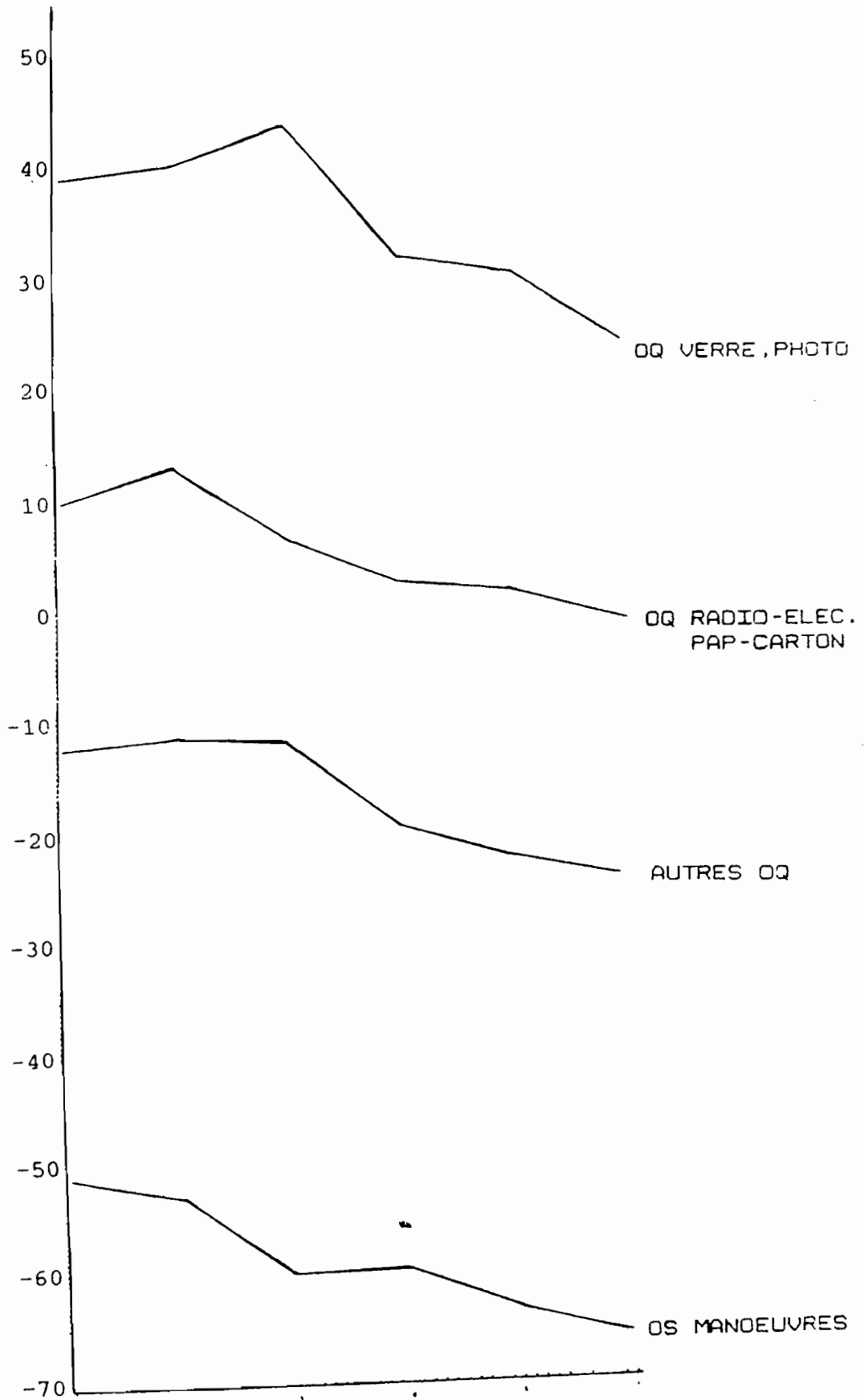
NOTES

- [1] BOUILLAGUET P. & all : Femmes au travail , Prospérité et crise Economica, 1981 p. 119.
- [2] L'information discriminante symétrisée est telle que l'information discriminante entre deux tableaux T et T' est identique à celle entre T' et T ($N(T/T')=N(T'/T)$).
- [3] Voir formulation en Annexe A.9
- [4] TURPIN E., MARY S.: Panorama économique des régions françaises. Collec. de l'INSEE série R, n° 42/43 Janv. 1981. p. 136.
- [5] GOFFINET P.: Les marchés locaux du travail en Rhône-Alpes, Eco & Stat n°182 Nov 1985 , p.65.
- [6] Cet indice, qui croît avec le degré de diversité de l'industrie régionale, a été construit à partir de la structure de référence qui est la décomposition de l'emploi dans l'industrie en 52 branches (nomenclature NAP 100) La "distance" de chaque région est calculée par rapport à ce profil.
- [7] INSEE : Disparités et diversité des régions françaises à la veille du 9^{ème} plan , Archives et Documents n°109. p.142.
- [8] MAURAU G. : Localisation de l'emploi du tertiaire et crise économique in L'emploi du tertiaire , Economica 1985. p. 78.
- [9] MAURAU G. : Op. cité p. 76.
- [10] BRAIBANT M. : Le tertiaire insaisissable ? Eco & Stat n° 146 Juillet Août 1982 p. 5.
- [11] LIPIETZ A. : La dimension régionale du développement du tertiaire, Cahier du CEPREMAP 1978 p. 8.
- [12] MARC N. , MARCHAND O. : La population active de 1975 à 1982 : les facteurs d'une forte croissance. Eco & Stat n° 171/172 p. 14.
- [13] ZIGHERA J.A. , GUILLOU A.: Etude d'un effet régional sur la participation féminine à l'emploi. CASSF-Paris X-Nanterre, Rapport pour le Ministère du plan. Année 1982.
- [14] HUET M. : Concentration des emplois féminins, Eco & Stat n° 154 p. 36.
- [15] BOUILLAGUET B. & all : op. cité p. 157.
- [16] HUET M. : Op. cité p. 36.
- [17] JAYET H. , JEANNIC T. Croissance démographique régionale les disparités s'atténuent légèrement n° 190 p. 25.

- [18] MICHAL M.G. : L'emploi féminin en 1968 . Collections de l'INSEE serie D n° 25 Nov. 1973. Voir également "Panorama économique des régions françaises" Op. cité.
- [19] GIRAN J.P, GARNIER R. : Structures par âge du chômage (définition BIT) moyennes annuelles in Politique de l'emploi , Economica 1983 p. 25.
- [20] DIRECTION DE LA PREVISION : Vers une précarisation accrue des jeunes sur le marché du travail ? Mai 1983 p. 2.
- [21] Cité par DRUMAN B. : Insertion des jeunes : efficacité et limite des politiques de l'emploi,colloque structures du marché du travail et politiques de l'emploi,3-4 Oct. 1985.
- [22] HUET M. : La progression de l'activité féminine est-elle irréversible, Eco & Stat n° 145 JUIN 1982 p. 13.
- [23] VINCENS J. : Problématique générale de l'insertion professionnelle communication colloque Louvain 1981.
- [24] INSEE : Les résultats de l'enquête emploi d'Octobre 1979, Eco & Stat n° 121 1983 p. 23.
- [25] MALGAIVE M. : Fonctionnement du dispositif et stratégie des jeunes Rapport CNAM , Fev. 1985.
- [26] ENQUETE CEREQ : Itinéraire dans le dispositif de formation et situation à la sortie de stages, Formation et Emploi , Janv/Mars 1985 p. 59.
- [27] DRUGMAN B., MOUY P.: Insertion des jeunes : efficacité et limite des politiques de l'emploi, Op. cité p. 12.
- [28] AGLIETTA M. : Panorama sur les théories de l'emploi, Revue Economique 1978, p. 116-117.
- [29] DURR G.M., SABOULIN M. : L'âge des Français : contrastes régionaux et opposition ville-campagne , Eco & Stat n°173, Janv. 1985 p. 26
- [30] FRANCCART G. : Le rééquilibrage démographique de la France, Eco & Stat n° 153 Mars 1983 p. 38.
- [31] GUIOT B.,HERANO.: Les jeunes et leur région , in Archives et Documents n° 230, INSEE Fev. 88.
- [32] INSEE : Disparités et diversité des régions françaises à la veille du 9^{ème} plan, Archives et Documents n°109 juillet 1984.
- [33] HANNOUN M., SICHERMAN : Résorption des disparités régionales et nouveaux clivages, Eco & Stat n° 153 Mars 83.
- [34] AUBRY B. : Les marchés locaux du travail en Alsace. Eco & Stat n° 182 Nov. 1985 p. 46.

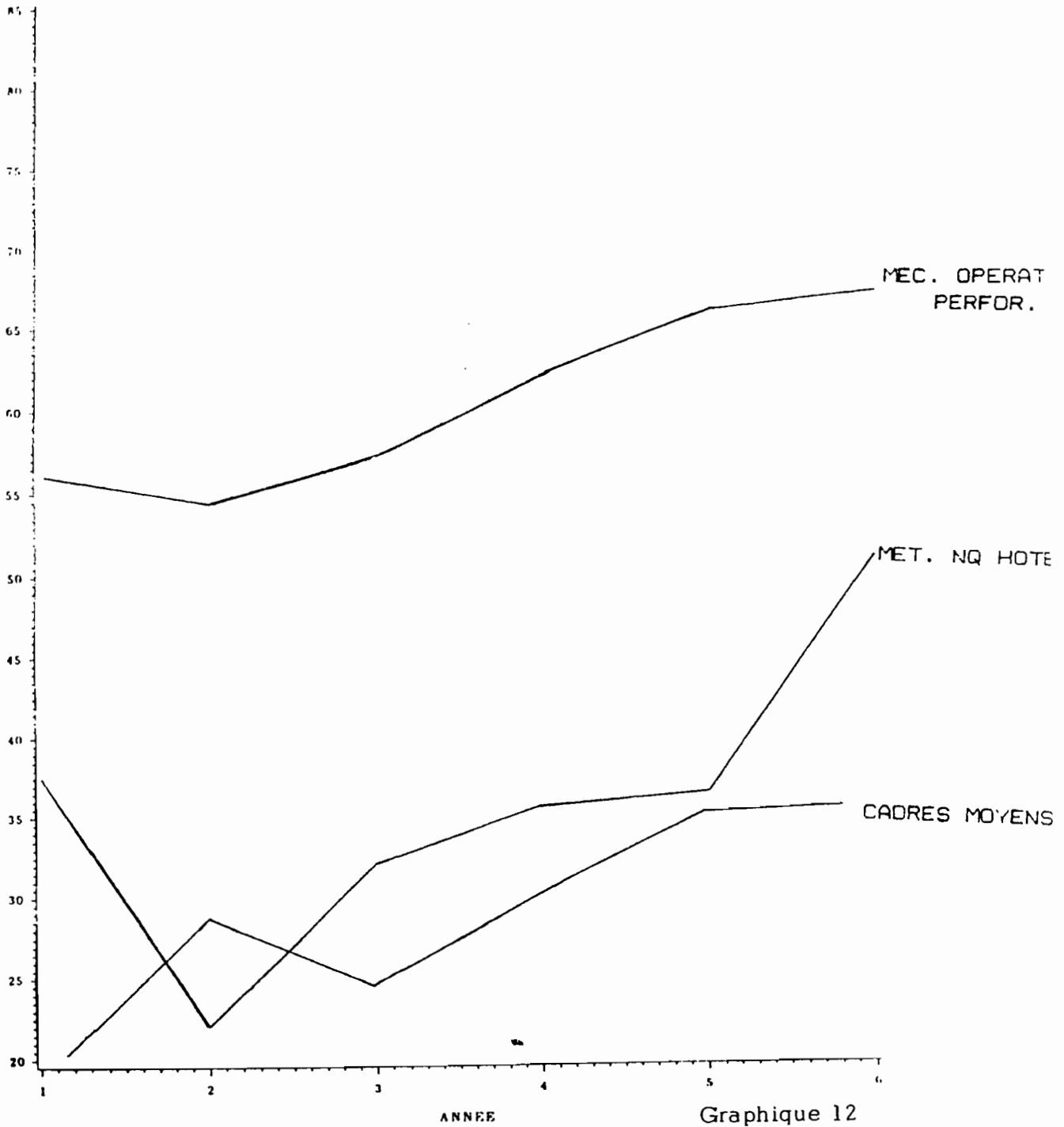
ANNEXE

EVOLUTION DE L'EMPLOI PAR PROFESS. EN IDF (1976-1981)

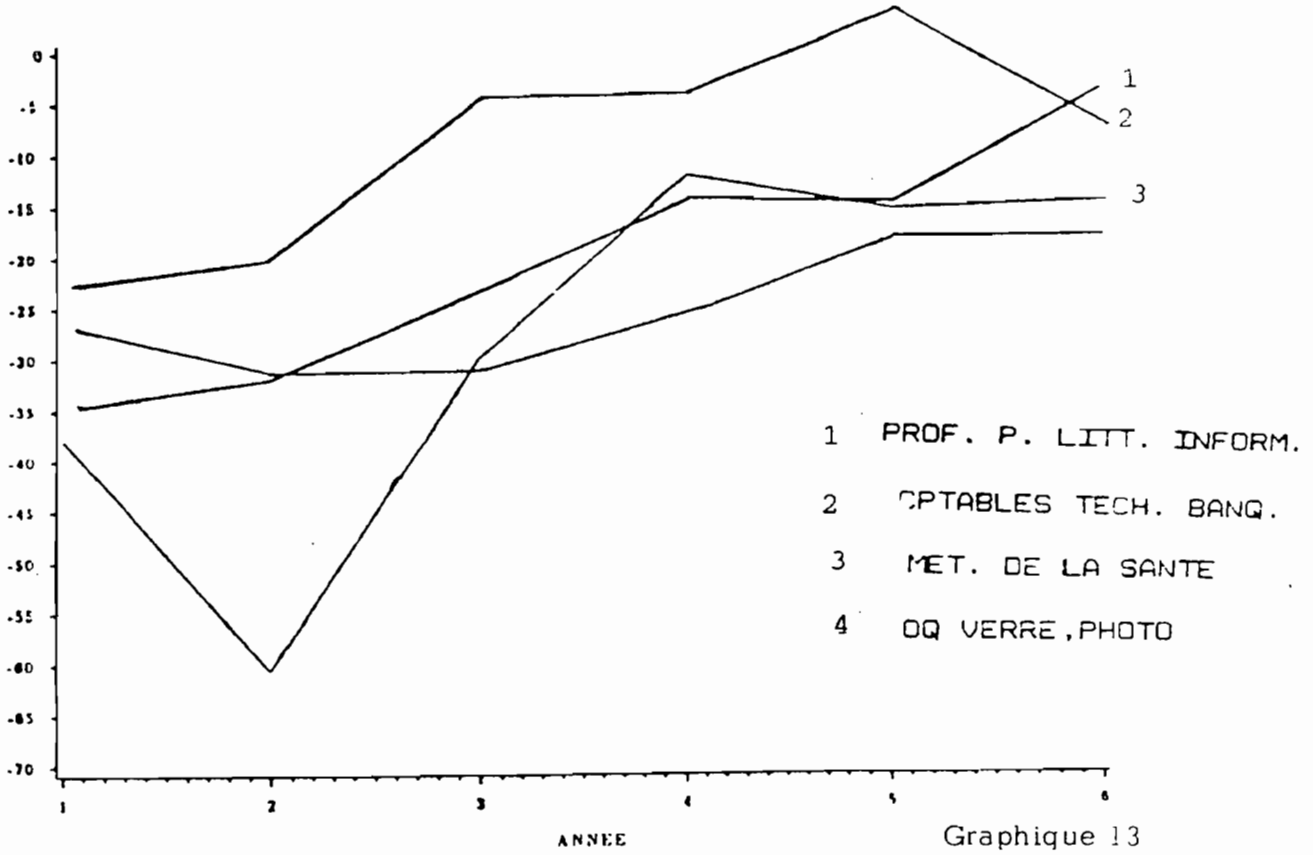


Graphique 11

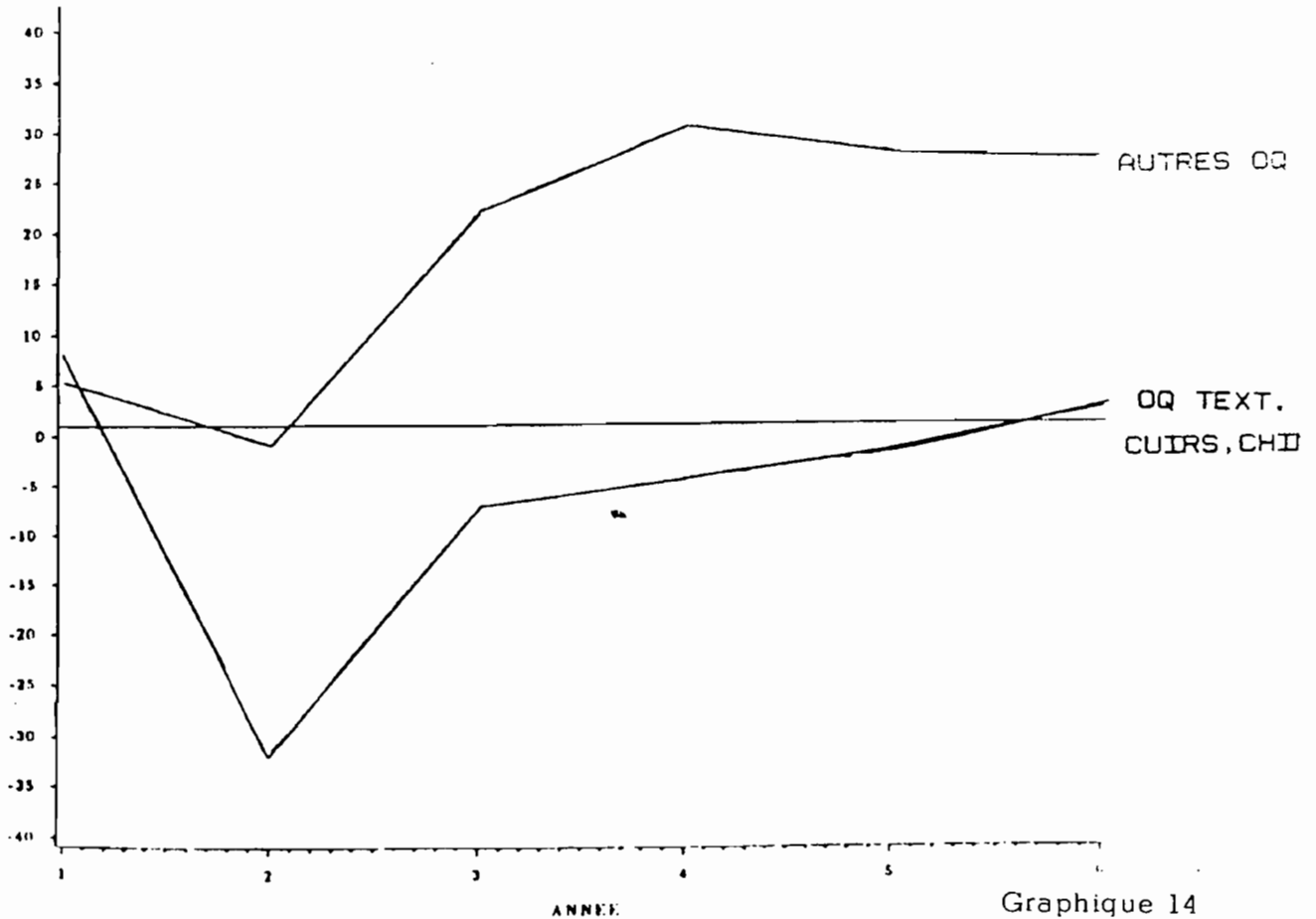
EVOLUTION DE L'EMPLOI PAR PROFESS. EN IDF (1976-1981)



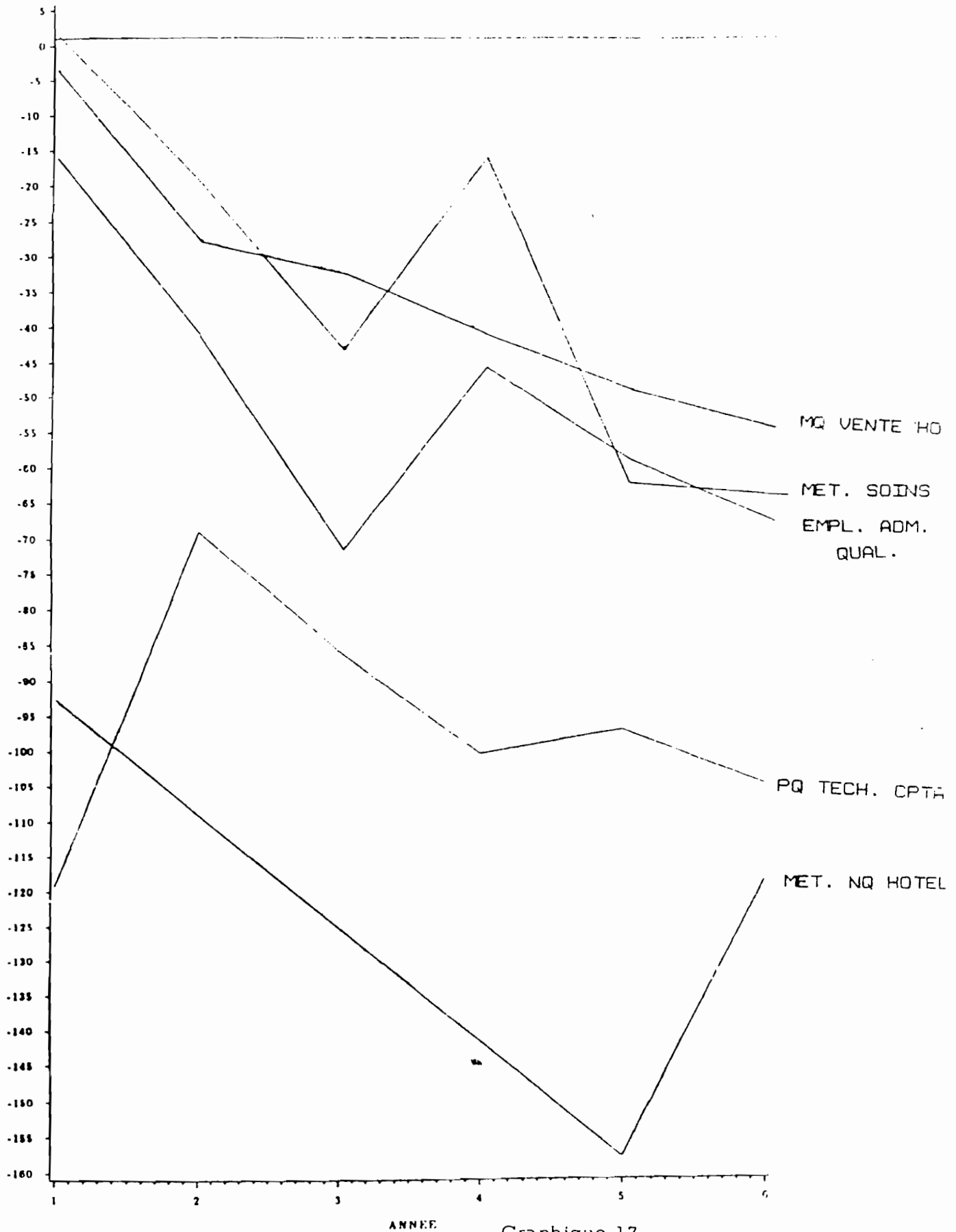
EVOLUTION DE L'EMPLOI PAR PROFESS. EN RHA (1976-1981)



EVOLUTION DE L'EMPLOI PAR PROFESS. EN FCOM (1976-1981)



EVOLUTION DE L'EMPLOI PAR PROFESS. EN FCOM (1976-1981)



Graphique 17

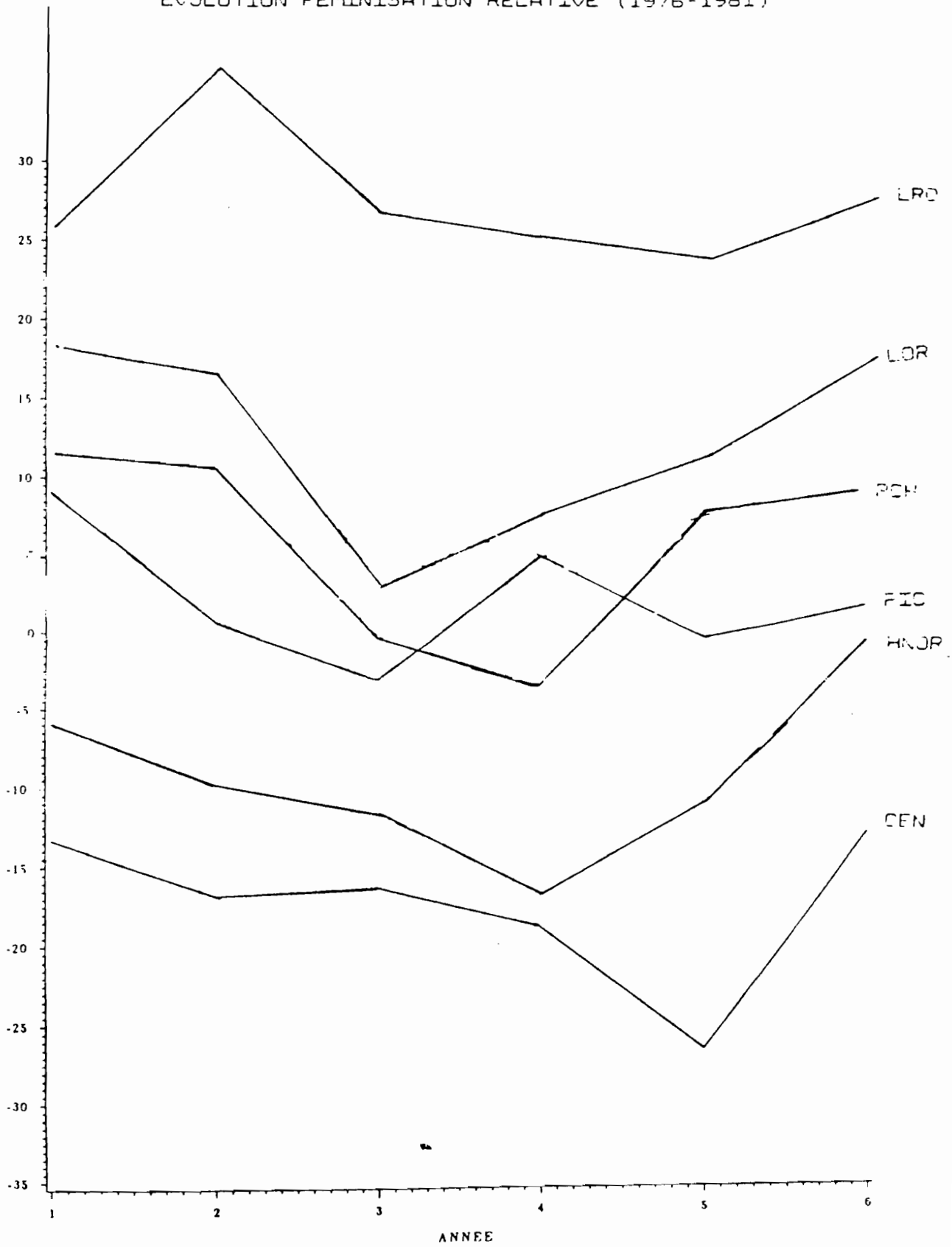
| | | | | | | | | | | | | | | | |
|--------|-------|-------|-------|--------|--------|-------|-------|--------|-------|-------|-------|-------|-------|-------|-------|
| 234. | -336. | -8. | -24. | 378. | 664. | -43. | 43. | 523. | -222. | 353. | 422. | 364. | 294. | 433. | -120. |
| -642. | 17. | 32. | 591. | 308. | 273. | -221. | 72. | 1461. | 753. | 1316. | 1390. | 1257. | 1261. | 1421. | 748. |
| 1118. | 663. | 928. | 899. | 1322. | 1596. | 900. | 997. | 2. | 15. | 12. | 8. | 3. | 10. | 17. | 2. |
| 483. | 979. | 987. | 1592. | 1238. | 1235. | 763. | 1035. | 2. | 15. | 12. | 8. | 3. | 10. | 17. | 2. |
| 3. | 5. | 6. | 2. | 5. | 3. | 3. | 12. | 2. | 15. | 12. | 8. | 3. | 10. | 17. | 2. |
| 24. | 8. | 7. | 22. | 2. | 18. | 35. | 226. | 2. | 15. | 12. | 8. | 3. | 10. | 17. | 2. |
| 552. | -27. | 13. | -113. | -236. | -659. | 151. | -71. | -474. | -89. | 27. | -233. | -415. | 221. | -107. | 379. |
| 296. | -98. | -286. | -216. | -631. | -286. | -121. | 26. | 564. | 901. | 994. | 757. | 604. | 1227. | 867. | 1425. |
| 1610. | 945. | 892. | 861. | 749. | 445. | 1145. | 931. | 0. | 2. | 1. | 1. | 0. | 1. | 1. | 0. |
| 1293. | 914. | 752. | 743. | 507. | 739. | 883. | 988. | 0. | 2. | 1. | 1. | 0. | 1. | 1. | 0. |
| 1. | 1. | 1. | 0. | 0. | 0. | 0. | 1. | 0. | 2. | 1. | 1. | 0. | 1. | 1. | 0. |
| 8. | 1. | 1. | 1. | 0. | 1. | 5. | 28. | 0. | 2. | 1. | 1. | 0. | 1. | 1. | 0. |
| 220. | -115. | 37. | 76. | -409. | -707. | -499. | 254. | -1165. | -390. | -647. | -206. | -418. | -195. | -287. | 333. |
| 324. | 17. | -215. | -669. | -94. | -99. | 112. | 45. | 278. | 655. | 498. | 763. | 591. | 795. | 711. | 1337. |
| 1133. | 850. | 997. | 1021. | 619. | 417. | 587. | 1255. | 0. | 2. | 1. | 1. | 0. | 1. | 1. | 0. |
| 1306. | 1006. | 792. | 464. | 851. | 875. | 1094. | 1006. | 0. | 2. | 1. | 1. | 0. | 1. | 1. | 0. |
| 0. | 1. | 1. | 0. | 0. | 0. | 0. | 2. | 0. | 2. | 1. | 1. | 0. | 1. | 1. | 0. |
| 9. | 1. | 1. | 1. | 0. | 2. | 7. | 31. | 0. | 2. | 1. | 1. | 0. | 1. | 1. | 0. |
| -1581. | -125. | -151. | -875. | -334. | -1115. | 192. | -62. | -380. | 33. | -640. | -404. | -305. | -598. | -437. | 222. |
| 143. | -225. | -306. | -508. | 537. | 135. | 370. | 56. | 602. | 989. | 496. | 619. | 654. | 525. | 605. | 1183. |
| 185. | 832. | 818. | 390. | 660. | 274. | 1159. | 912. | 0. | 3. | 1. | 1. | 0. | 1. | 1. | 0. |
| 1077. | 781. | 716. | 539. | 1582. | 1093. | 1401. | 1018. | 0. | 3. | 1. | 1. | 0. | 1. | 1. | 0. |
| 0. | 1. | 1. | 0. | 0. | 0. | 1. | 2. | 0. | 3. | 1. | 1. | 0. | 1. | 1. | 0. |
| 8. | 1. | 1. | 1. | 0. | 2. | 9. | 34. | 0. | 3. | 1. | 1. | 0. | 1. | 1. | 0. |
| 279. | -142. | -448. | 149. | -515. | -693. | -550. | -183. | -609. | -311. | -52. | -63. | -273. | -394. | -299. | 101. |
| 80. | 23. | -59. | -109. | 190. | 4. | 363. | 33. | 490. | 716. | 913. | 890. | 691. | 659. | 711. | 1072. |
| 1217. | 836. | 621. | 1111. | 563. | 427. | 564. | 827. | 0. | 3. | 2. | 1. | 0. | 1. | 2. | 1. |
| 1035. | 1024. | 937. | 822. | 1143. | 981. | 1422. | 995. | 0. | 3. | 2. | 1. | 0. | 1. | 2. | 1. |
| 1. | 1. | 1. | 0. | 0. | 0. | 0. | 2. | 0. | 3. | 2. | 1. | 0. | 1. | 2. | 1. |
| 9. | 1. | 1. | 2. | 0. | 2. | 11. | 42. | 0. | 3. | 2. | 1. | 0. | 1. | 2. | 1. |
| 179. | 621. | -175. | 63. | -1008. | -744. | -121. | 13. | -451. | -103. | -422. | 108. | 155. | -421. | -284. | -267. |
| 260. | -161. | -138. | -391. | -369. | -208. | 79. | 37. | 572. | 880. | 628. | 1053. | 1056. | 639. | 719. | 740. |
| 1097. | 1789. | 813. | 1016. | 343. | 405. | 864. | 1002. | 0. | 1. | 1. | 1. | 0. | 1. | 1. | 0. |
| 1234. | 849. | 863. | 618. | 652. | 791. | 1067. | 998. | 0. | 1. | 1. | 1. | 0. | 1. | 1. | 0. |
| 0. | 1. | 1. | 0. | 0. | 0. | 0. | 1. | 0. | 1. | 1. | 1. | 0. | 1. | 1. | 0. |
| 6. | 1. | 1. | 1. | 0. | 1. | 5. | 22. | 0. | 1. | 1. | 1. | 0. | 1. | 1. | 0. |
| -436. | 37. | -285. | -339. | -17. | -236. | -116. | -111. | -770. | 61. | 188. | -52. | 366. | -93. | -139. | 7. |
| 154. | 34. | -154. | -334. | -117. | -71. | 46. | 12. | 426. | 1062. | 1184. | 920. | 1338. | 909. | 852. | 997. |
| 608. | 1022. | 747. | 697. | 947. | 690. | 890. | 907. | 0. | 2. | 1. | 1. | 0. | 1. | 1. | 0. |
| 1138. | 1058. | 871. | 671. | 859. | 930. | 1058. | 974. | 0. | 2. | 1. | 1. | 0. | 1. | 1. | 0. |
| 0. | 1. | 1. | 0. | 0. | 0. | 0. | 1. | 0. | 2. | 1. | 1. | 0. | 1. | 1. | 0. |
| 6. | 1. | 1. | 1. | 0. | 1. | 5. | 26. | 0. | 2. | 1. | 1. | 0. | 1. | 1. | 0. |
| -809. | -32. | -253. | -457. | -109. | 205. | -136. | -185. | -833. | 126. | -220. | -475. | -230. | -62. | -239. | -38. |
| 302. | -122. | 165. | -329. | 85. | -313. | 30. | 28. | 394. | 1115. | 776. | 593. | 725. | 923. | 759. | 938. |
| 412. | 939. | 759. | 610. | 849. | 1054. | 858. | 829. | 0. | 6. | 2. | 1. | 0. | 1. | 3. | 1. |
| 1298. | 891. | 1179. | 663. | 1035. | 719. | 1024. | 990. | 0. | 6. | 2. | 1. | 0. | 1. | 3. | 1. |
| 0. | 2. | 1. | 0. | 1. | 1. | 1. | 3. | 0. | 6. | 2. | 1. | 0. | 1. | 3. | 1. |
| 19. | 2. | 3. | 3. | 1. | 3. | 14. | 68. | 0. | 6. | 2. | 1. | 0. | 1. | 3. | 1. |

| | | | | | | | | | | | | | | |
|-------|-------|-------|-------|--------|-------|-------|---------|-------|-------|--------|--------|-------|-------|-------|
| 255. | 371. | 104. | -408. | -499. | 60. | -143. | -539. | -76. | 31. | -503. | -982. | 21. | -336. | 69. |
| 106. | 120. | 82. | -406. | -114. | 106. | 19. | -114. | 106. | 19. | 1052. | 873. | 1012. | 694. | 1053. |
| 1204. | 1418. | 1094. | 646. | 526. | 1052. | 873. | 533. | 920. | 1006. | 581. | 345. | 1012. | 694. | 1053. |
| 1077. | 1145. | 1094. | 619. | 884. | 1116. | 981. | 0. | 3. | 1. | 1. | 0. | 1. | 1. | 0. |
| 1. | 2. | 1. | 0. | 0. | 1. | 2. | 0. | 3. | 1. | 1. | 0. | 1. | 1. | 0. |
| 9. | 2. | 1. | 1. | 2. | 9. | 39. | | | | | | | | |
| 66. | -114. | -488. | 74. | -372. | 397. | 445. | -137. | -21. | -663. | -287. | 412. | 36. | -344. | -257. |
| 328. | 57. | -241. | -531. | -218. | -112. | 43. | | | | | | | | |
| 973. | 853. | 591. | 1021. | 584. | 1441. | 1534. | 778. | 949. | 491. | 705. | 1358. | 1003. | 673. | 742. |
| 1314. | 1050. | 774. | 534. | 779. | 876. | 1004. | 0. | 2. | 1. | 1. | 0. | 1. | 1. | 0. |
| 0. | 1. | 1. | 0. | 0. | 1. | 2. | 0. | 2. | 1. | 1. | 0. | 1. | 1. | 0. |
| 9. | 1. | 1. | 1. | 2. | 6. | 31. | | | | | | | | |
| -811. | -299. | -656. | -624. | -1235. | -437. | 25. | -749. | -494. | -711. | -1372. | -1578. | -635. | -587. | -240. |
| 486. | -14. | -478. | -270. | 179. | 278. | 107. | 395. | 555. | 438. | 223. | 174. | 481. | 495. | 708. |
| 380. | 664. | 469. | 477. | 231. | 587. | 946. | 0. | 1. | 0. | 0. | 0. | 0. | 1. | 0. |
| 1443. | 917. | 572. | 650. | 1086. | 1214. | 1071. | 0. | 1. | 0. | 0. | 0. | 0. | 1. | 0. |
| 0. | 1. | 0. | 0. | 0. | 0. | 1. | | | | | | | | |
| 7. | 1. | 0. | 1. | 2. | 5. | 23. | | | | | | | | |
| -697. | -48. | -268. | -27. | -341. | -164. | -21. | -40. | 15. | -296. | -216. | -814. | 12. | -26. | -615. |
| 124. | -133. | 124. | -27. | -93. | 142. | 14. | | | | | | | | |
| 468. | 938. | 758. | 951. | 620. | 846. | 992. | 883. | 1012. | 729. | 779. | 411. | 1009. | 952. | 535. |
| 1103. | 894. | 1137. | 910. | 908. | 1163. | 975. | 0. | 4. | 1. | 1. | 0. | 2. | 2. | 0. |
| 0. | 2. | 1. | 0. | 0. | 1. | 3. | | | | | | | | |
| 12. | 1. | 2. | 2. | 3. | 11. | 51. | | | | | | | | |
| -428. | 623. | 54. | -322. | -1433. | -249. | -249. | -146. | 281. | -247. | 102. | 150. | -13. | -138. | -159. |
| -19. | 102. | 112. | -19. | -203. | 59. | 28. | | | | | | | | |
| 603. | 1809. | 1031. | 698. | 205. | 1042. | 778. | 782. | 1302. | 755. | 1056. | 1061. | 969. | 839. | 831. |
| 942. | 1114. | 1117. | 752. | 810. | 1070. | 990. | 0. | 4. | 1. | 1. | 0. | 1. | 2. | 0. |
| 0. | 2. | 1. | 0. | 0. | 1. | 2. | | | | | | | | |
| 8. | 1. | 1. | 2. | 2. | 8. | 39. | | | | | | | | |
| -350. | 304. | -281. | 461. | -1251. | -6. | 173. | -82. | 319. | -220. | -523. | 132. | -126. | -234. | -919. |
| 244. | -140. | 93. | -576. | -349. | -85. | 38. | | | | | | | | |
| 646. | 1301. | 731. | 646. | 243. | 967. | 1175. | 826. | 1339. | 767. | 559. | 1031. | 858. | 755. | 385. |
| 1214. | 866. | 1085. | 513. | 686. | 904. | 999. | 0. | 3. | 1. | 0. | 0. | 1. | 1. | 0. |
| 0. | 1. | 1. | 0. | 0. | 0. | 1. | | | | | | | | |
| 7. | 1. | 1. | 1. | 1. | 5. | 25. | | | | | | | | |
| 18. | 167. | 182. | 465. | 46. | 123. | -228. | 158. | 164. | 17. | -295. | 357. | -118. | -28. | 64. |
| 43. | -116. | 141. | -245. | -56. | -21. | 10. | 1080. | 1180. | 1001. | 723. | 1328. | 889. | 954. | 1058. |
| 959. | 1167. | 1195. | 1561. | 916. | 1132. | 809. | 0. | 4. | 2. | 1. | 0. | 1. | 2. | 0. |
| 1021. | 912. | 1172. | 735. | 946. | 991. | 972. | 0. | 4. | 2. | 1. | 0. | 1. | 2. | 0. |
| 0. | 1. | 1. | 1. | 0. | 1. | 2. | | | | | | | | |
| 8. | 1. | 2. | 2. | 2. | 8. | 40. | | | | | | | | |
| 469. | 105. | 316. | 170. | -293. | 60. | 252. | 107. | 270. | 82. | -25. | 51. | 161. | -316. | 30. |
| 36. | -28. | -149. | -331. | 272. | -302. | 26. | | | | | | | | |
| 1482. | 1079. | 1344. | 1143. | 642. | 1046. | 793. | 1010. | 1291. | 1052. | 932. | 963. | 1156. | 704. | 1006. |
| 998. | 980. | 863. | 663. | 1293. | 737. | 987. | 0. | 4. | 1. | 1. | 0. | 1. | 1. | 0. |
| 1. | 1. | 1. | 0. | 0. | 0. | 2. | | | | | | | | |
| 8. | 1. | 1. | 2. | 3. | 6. | 37. | | | | | | | | |
| 254. | 438. | 135. | -348. | -14. | 413. | 189. | -18282. | 313. | 70. | -368. | -897. | -235. | -87. | -102. |

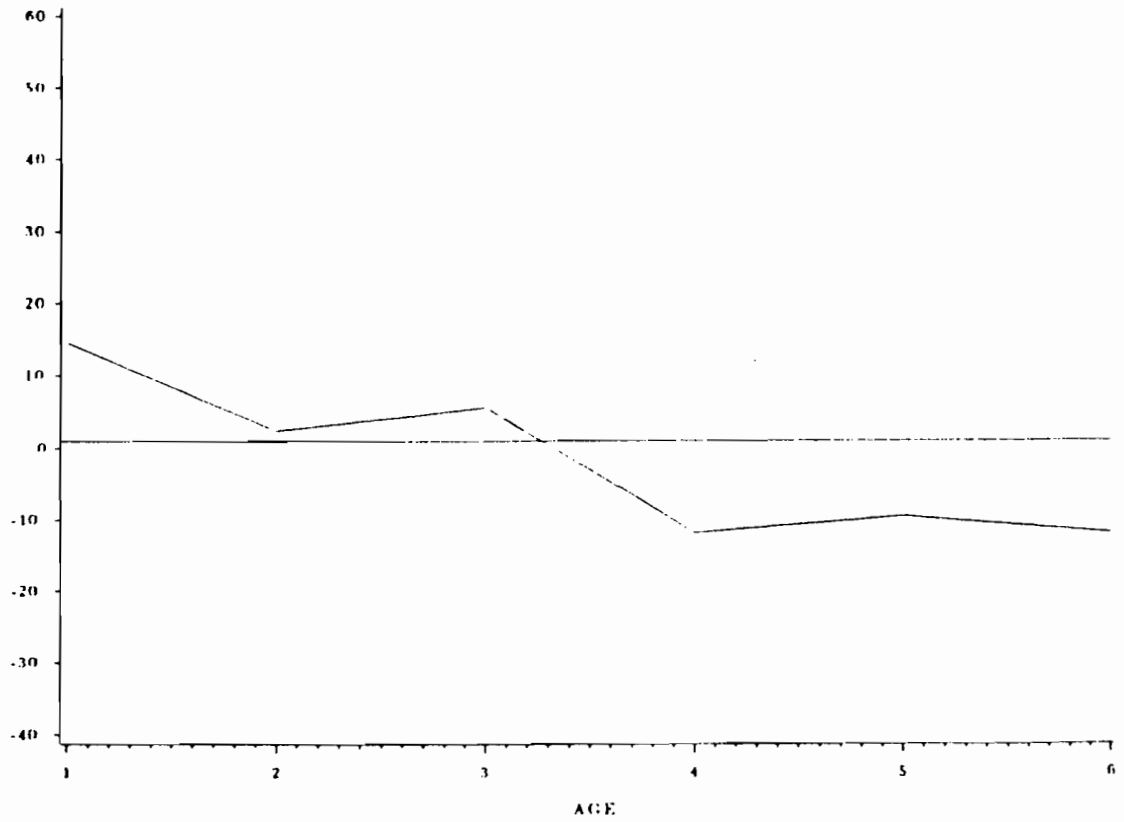
| | | | | | | |
|----------------------|----------------------|----------------------|---------------------|---------------------|---------------------|---------------------|
| -107.
895.
28. | -67.
932.
48. | 4.
1005.
48. | 61.
1062.
40. | 37.
1037.
32. | 99.
1095.
29. | 2.
997.
226. |
| 85.
1080.
5. | 117.
1117.
7. | -41.
956.
5. | -92.
907.
4. | 46.
1042.
4. | -263.
759.
2. | 7.
1002.
28. |
| 126.
1117.
6. | 163.
1161.
8. | 46.
1036.
6. | -125.
872.
5. | -151.
850.
4. | -364.
682.
2. | 13.
1008.
31. |
| 70.
1059.
6. | 158.
1158.
9. | 103.
1099.
8. | -167.
838.
5. | -249.
772.
4. | -127.
866.
3. | 11.
1007.
34. |
| -62.
938.
6. | 44.
1043.
10. | -1.
1001.
8. | -41.
961.
7. | 18.
1019.
6. | 34.
1027.
5. | 1.
996.
42. |
| 86.
1086.
4. | 15.
1013.
5. | -1.
1000.
4. | -28.
972.
4. | -120.
887.
3. | 25.
1017.
2. | 2.
997.
22. |
| -12.
986.
4. | 67.
1068.
6. | -19.
983.
5. | -67.
935.
4. | 40.
1041.
4. | -45.
950.
3. | 1.
996.
26. |
| 132.
1128.
13. | 182.
1187.
18. | -138.
865.
12. | -86.
910.
10. | -75.
920.
9. | -169.
831.
6. | 10.
1005.
68. |
| 135.
1136.
7. | 80.
1076.
9. | -75.
925.
7. | 8.
1003.
7. | -17.
978.
5. | -267.
757.
3. | 6.
1001.
39. |
| 42.
1037.
6. | 95.
1094.
8. | 6.
1004.
6. | 36.
1033.
5. | -101.
901.
4. | -236.
781.
3. | 5.
1000.
31. |
| -53.
944.
4. | 101.
1101.
6. | 93.
1096.
5. | -71.
930.
4. | -86.
916.
3. | -101.
895.
2. | 4.
999.
23. |
| 238.
1256.
10. | 71.
1064.
12. | -105.
895.
9. | -125.
876.
7. | -62.
933.
7. | -97.
894.
5. | 9.
1004.
51. |
| 60.
1058.
6. | -26.
972.
8. | 29.
1029.
8. | -122.
885.
6. | -23.
976.
5. | 106.
1103.
5. | 2.
998.
39. |

| | | | | | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|
| 1035. | 37. | -3. | -77. | 55. | 14. | -20. | 1. |
| 4. | 1035. | 995. | 927. | 1057. | 1015. | 973. | 996. |
| | | 5. | 5. | 4. | 4. | 3. | 25. |
| -83. | -83. | -183. | 52. | 25. | 94. | 169. | 7. |
| 913. | 913. | 827. | 1049. | 1020. | 1093. | 1170. | 1002. |
| 6. | 6. | 7. | 8. | 7. | 6. | 5. | 40. |
| -100. | -100. | -39. | -85. | 33. | 55. | 209. | 5. |
| 899. | 899. | 956. | 917. | 1030. | 1053. | 1220. | 1000. |
| 5. | 5. | 8. | 7. | 6. | 5. | 5. | 37. |
| -116. | -116. | -26. | 78. | -281. | 87. | 288. | 13. |
| 877. | 877. | 961. | 1070. | 746. | 1078. | 1309. | 1009. |
| 2. | 2. | 2. | 2. | 1. | 2. | 2. | 11. |
| 9. | 9. | -78. | 62. | 43. | 26. | -78. | 2. |
| 1006. | 1006. | 923. | 1085. | 1044. | 1027. | 918. | 997. |
| 16. | 16. | 22. | 23. | 18. | 15. | 11. | 105. |
| -174. | -174. | -144. | 125. | 69. | -60. | 211. | 10. |
| 832. | 832. | 857. | 1125. | 1063. | 935. | 1216. | 1005. |
| 3. | 3. | 4. | 5. | 4. | 3. | 3. | 23. |
| -27. | -27. | 5. | -22. | 120. | -49. | -63. | 2. |
| 970. | 970. | 1003. | 979. | 1127. | 952. | 932. | 997. |
| 4. | 4. | 6. | 6. | 5. | 4. | 3. | 28. |
| -157. | -157. | -139. | 20. | 88. | 134. | 99. | 6. |
| 848. | 848. | 864. | 1016. | 1086. | 1138. | 1091. | 1002. |
| 9. | 9. | 14. | 15. | 13. | 11. | 9. | 70. |
| 6. | 6. | 5. | 2. | 3. | 3. | 11. | 5. |
| 1001. | 1001. | 224. | 997. | 999. | 998. | 1006. | 1005. |
| 154. | 154. | | 202. | 167. | 139. | 114. | 1000. |

EVOLUTION FEMINISATION RELATIVE (1976-1981)

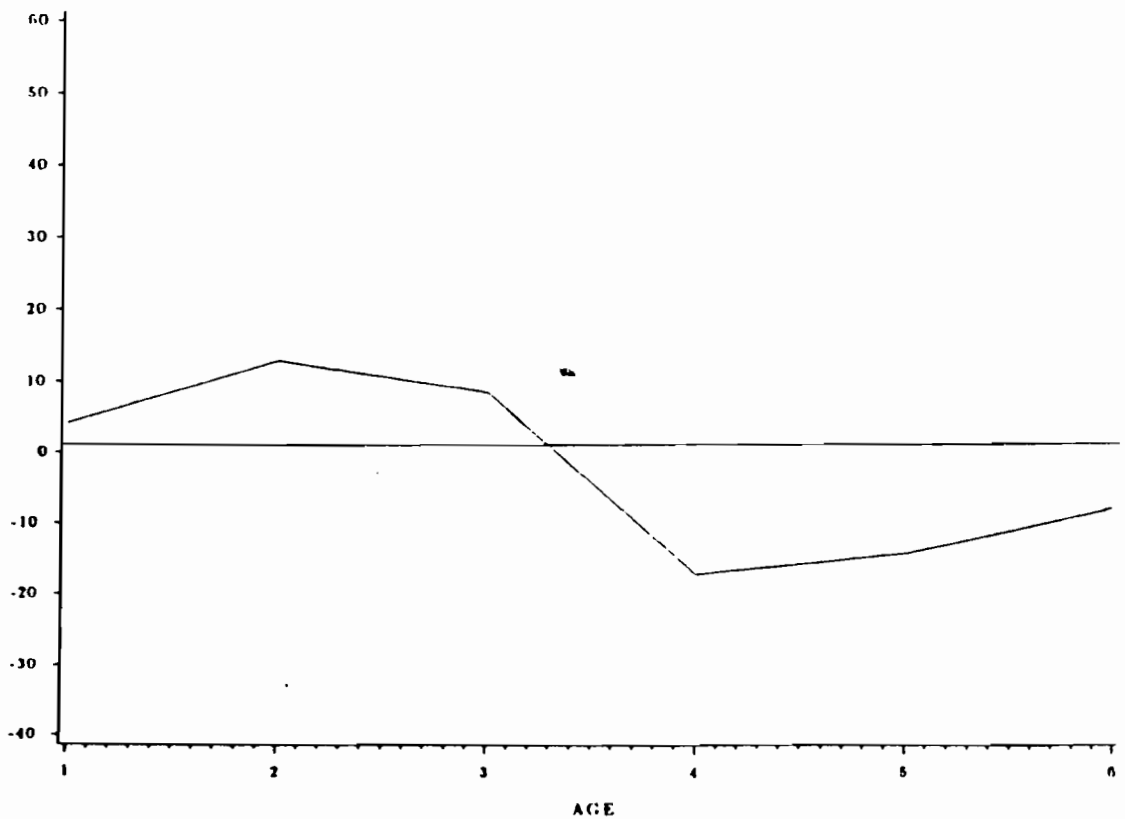


PROFIL REG. AGE

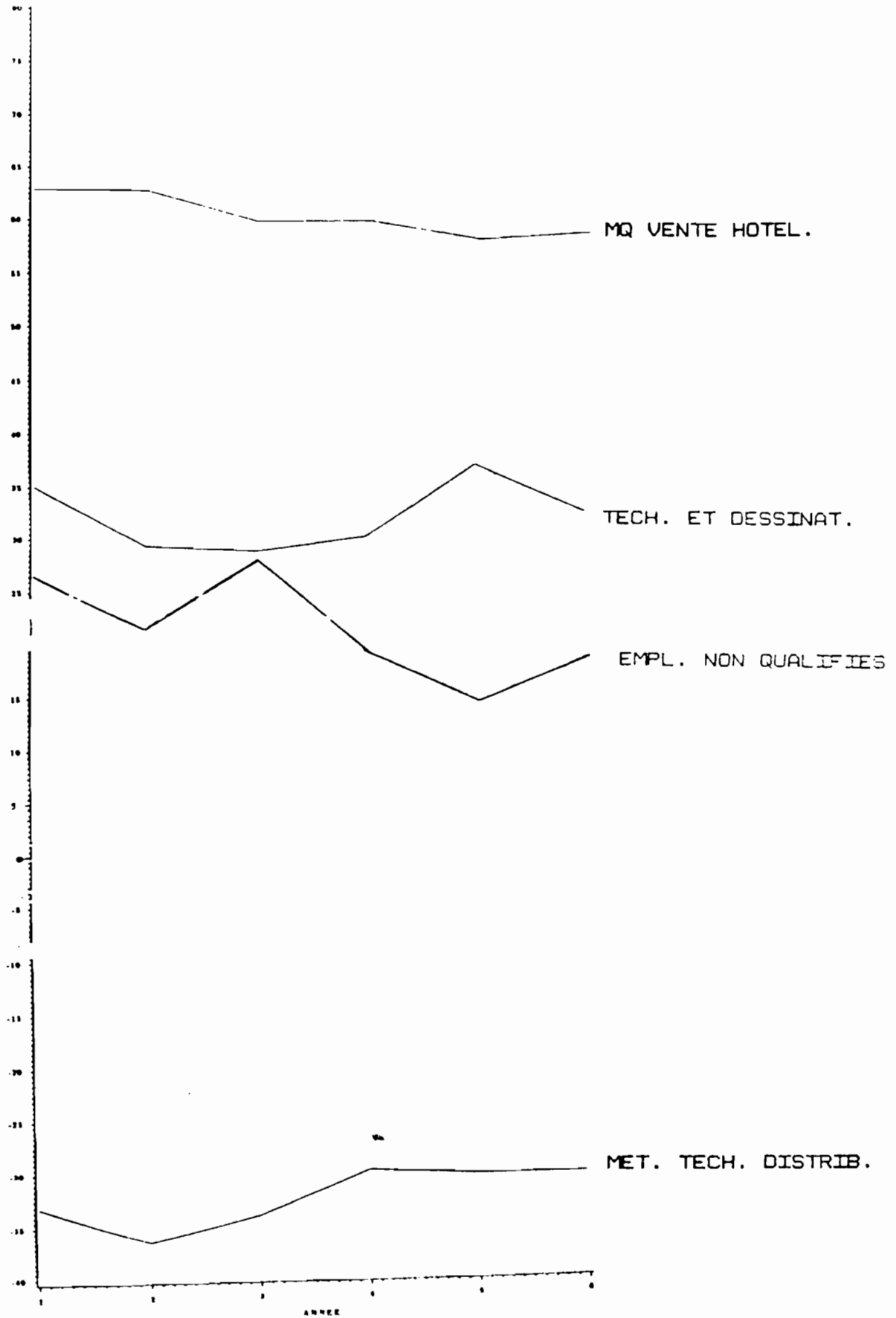


HAUTE NORMANDIE

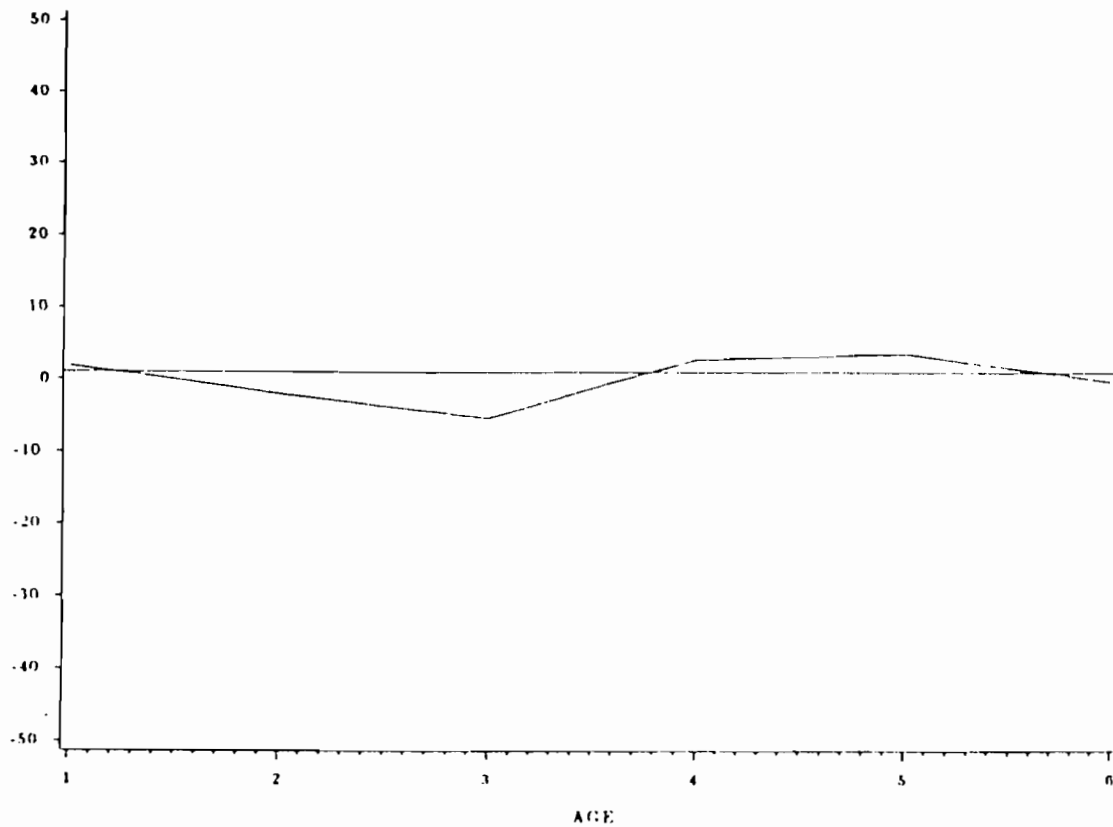
PROFIL REG. AGE



EVOLUTION EFFET PROFESSION (1976-1981)

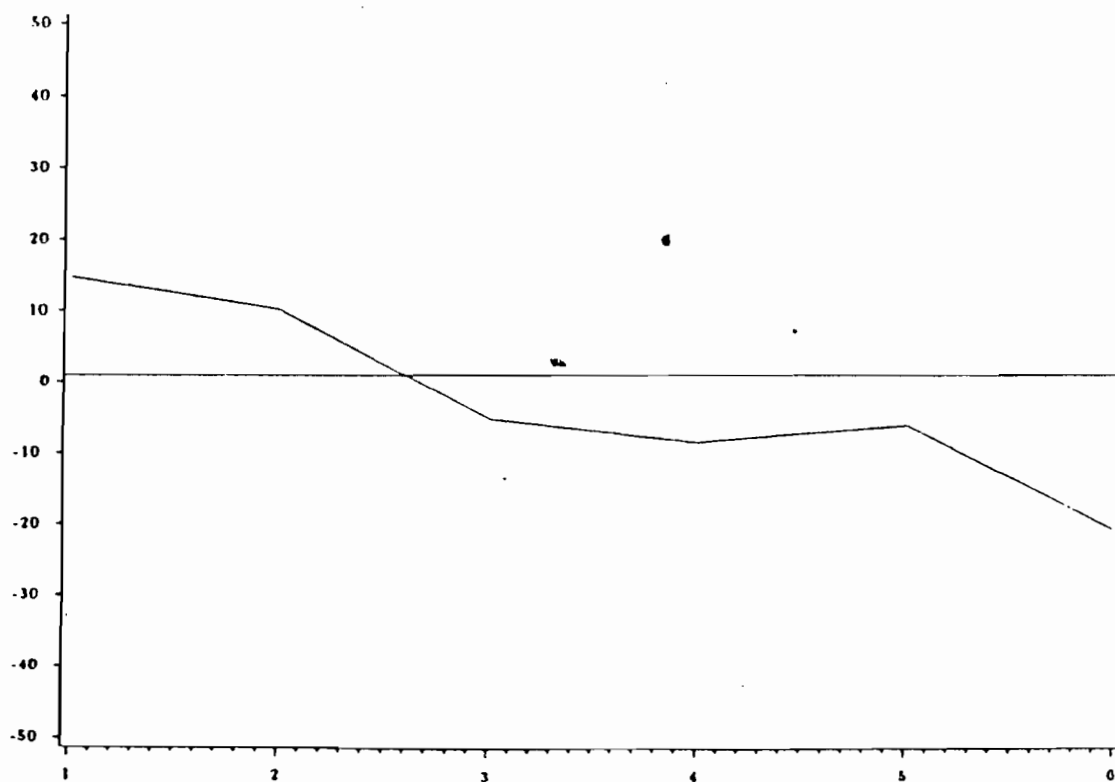


PROFIL REGAGE

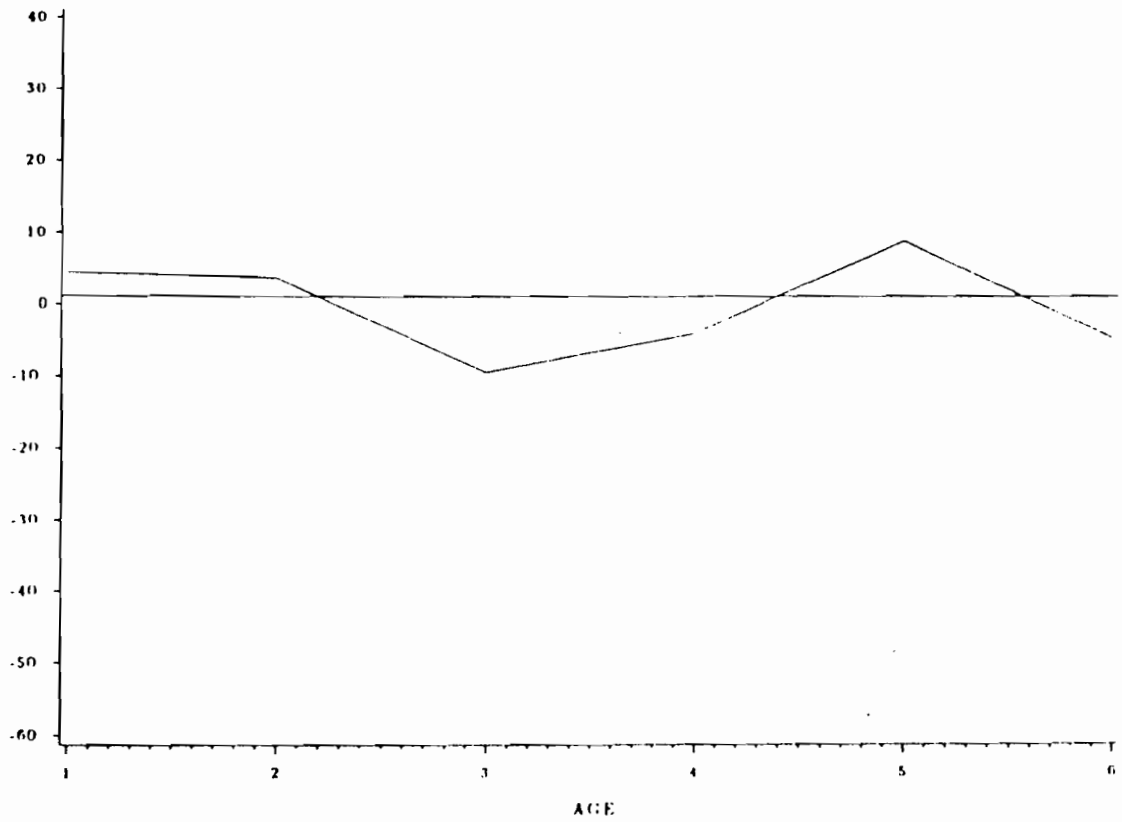


NORD-Pas-de CALAIS

PROFIL REGAGE

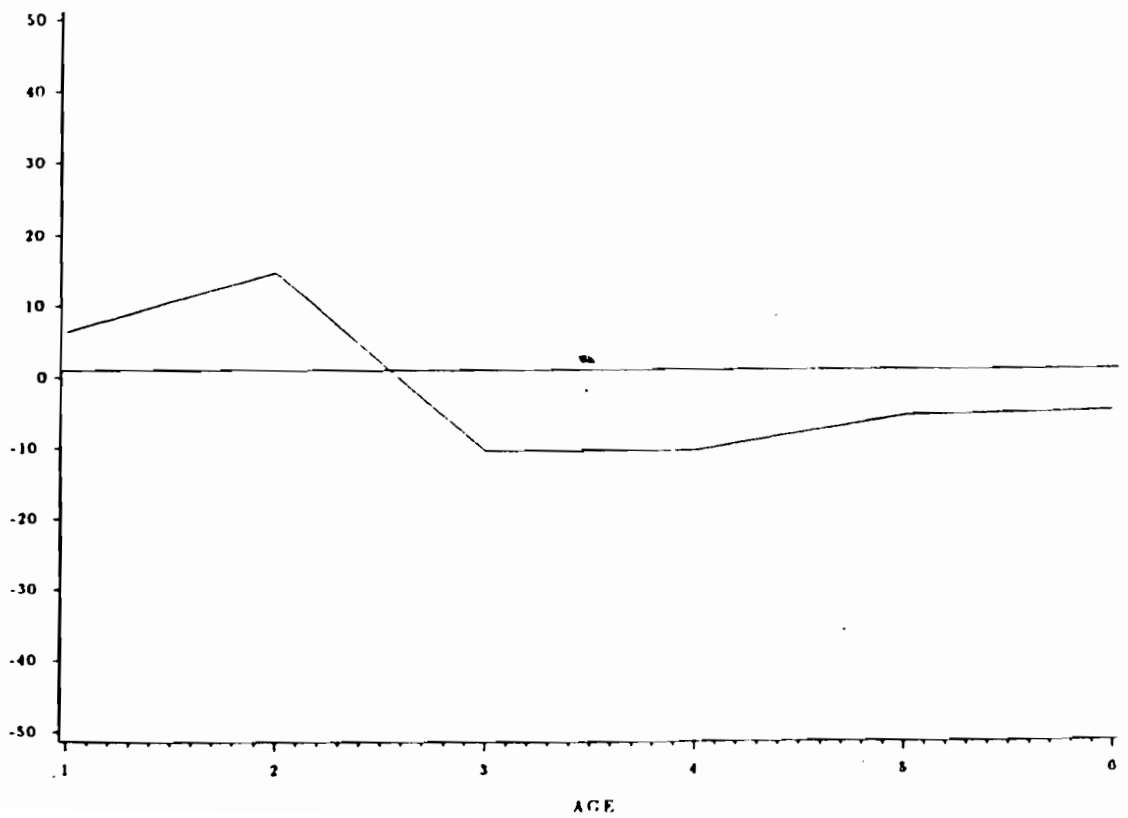


PROFIL REG. AGE

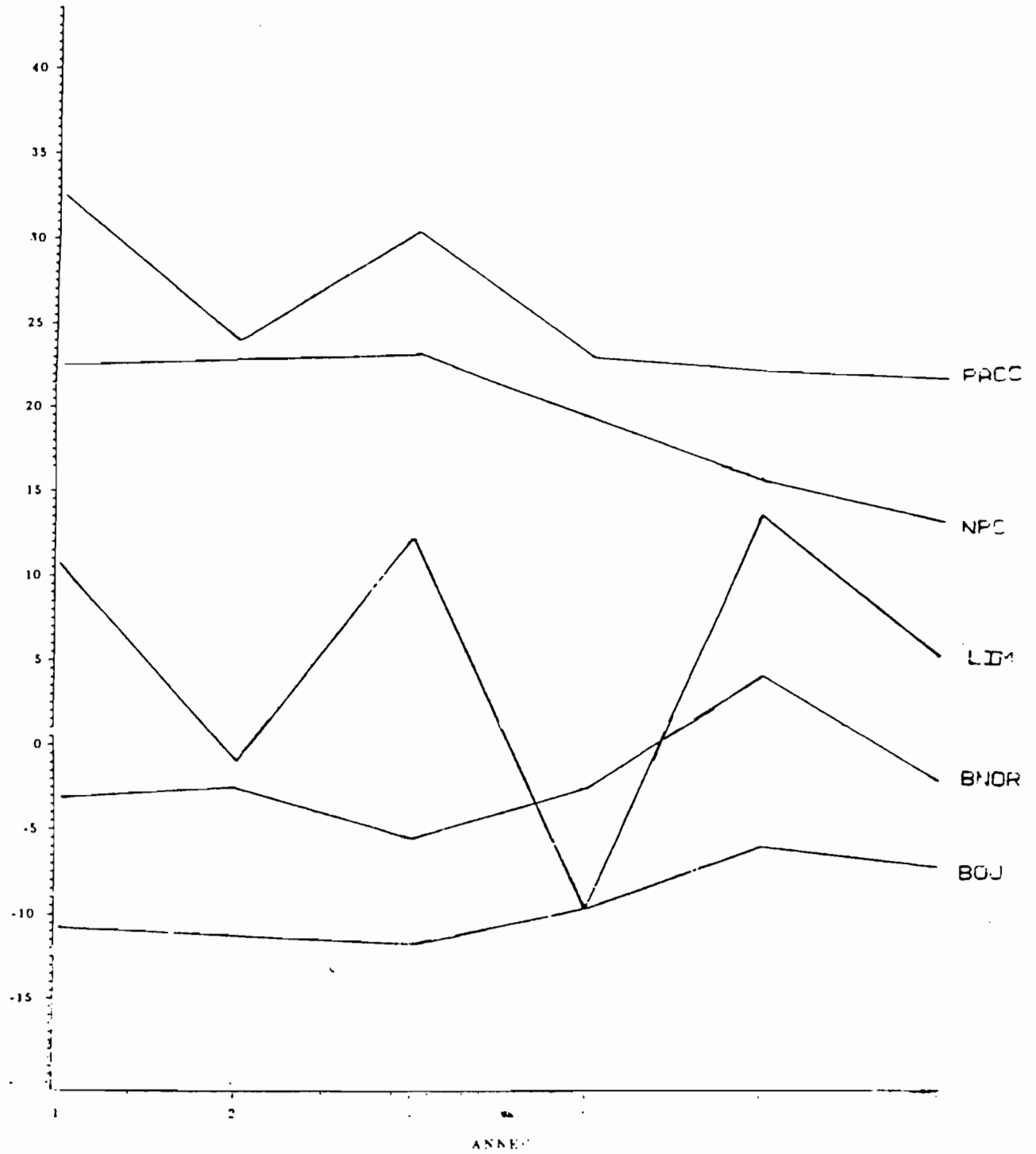


BASSE - NORMANDIE

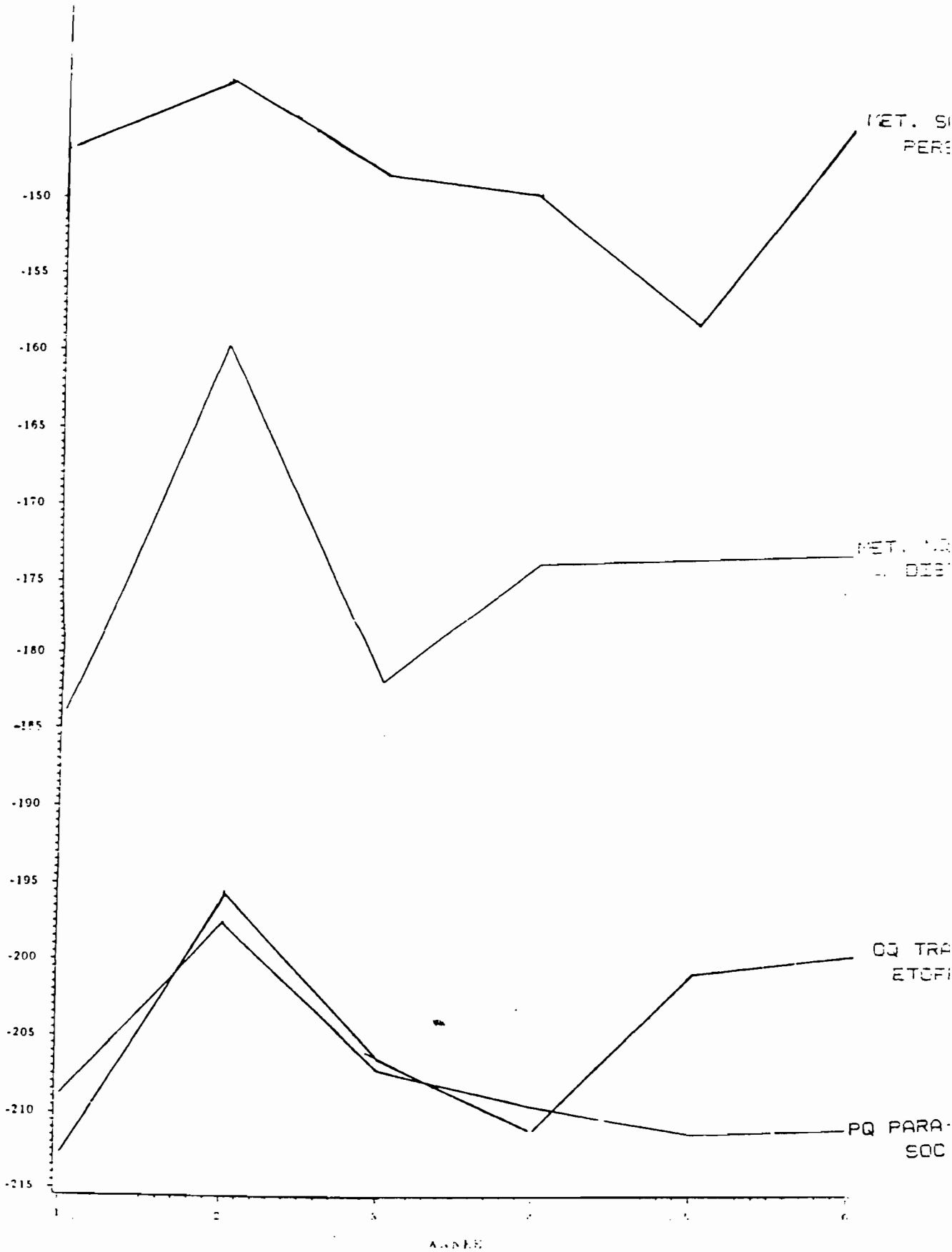
PROFIL REG. AGE



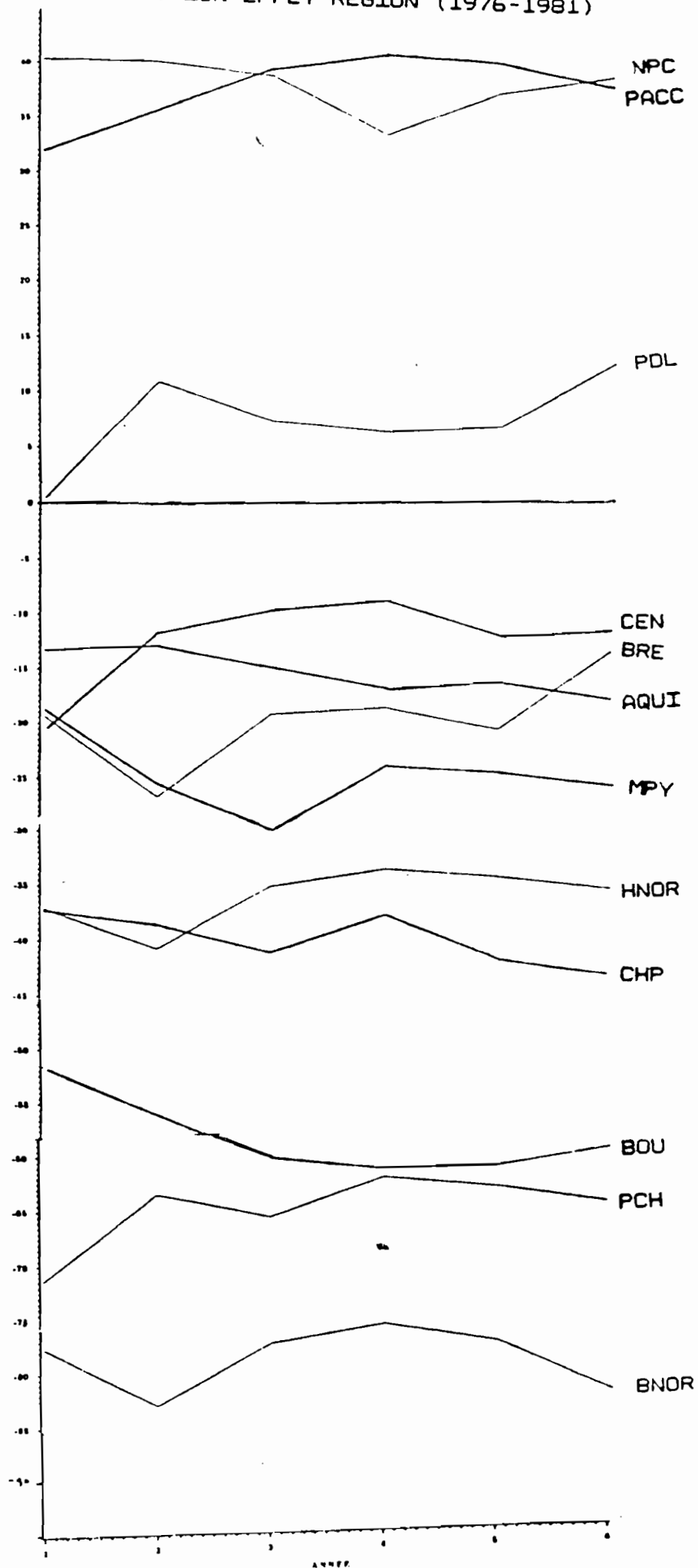
EVOLUTION FEMINISATION RELATIVE REGIONALE (1979-1991)



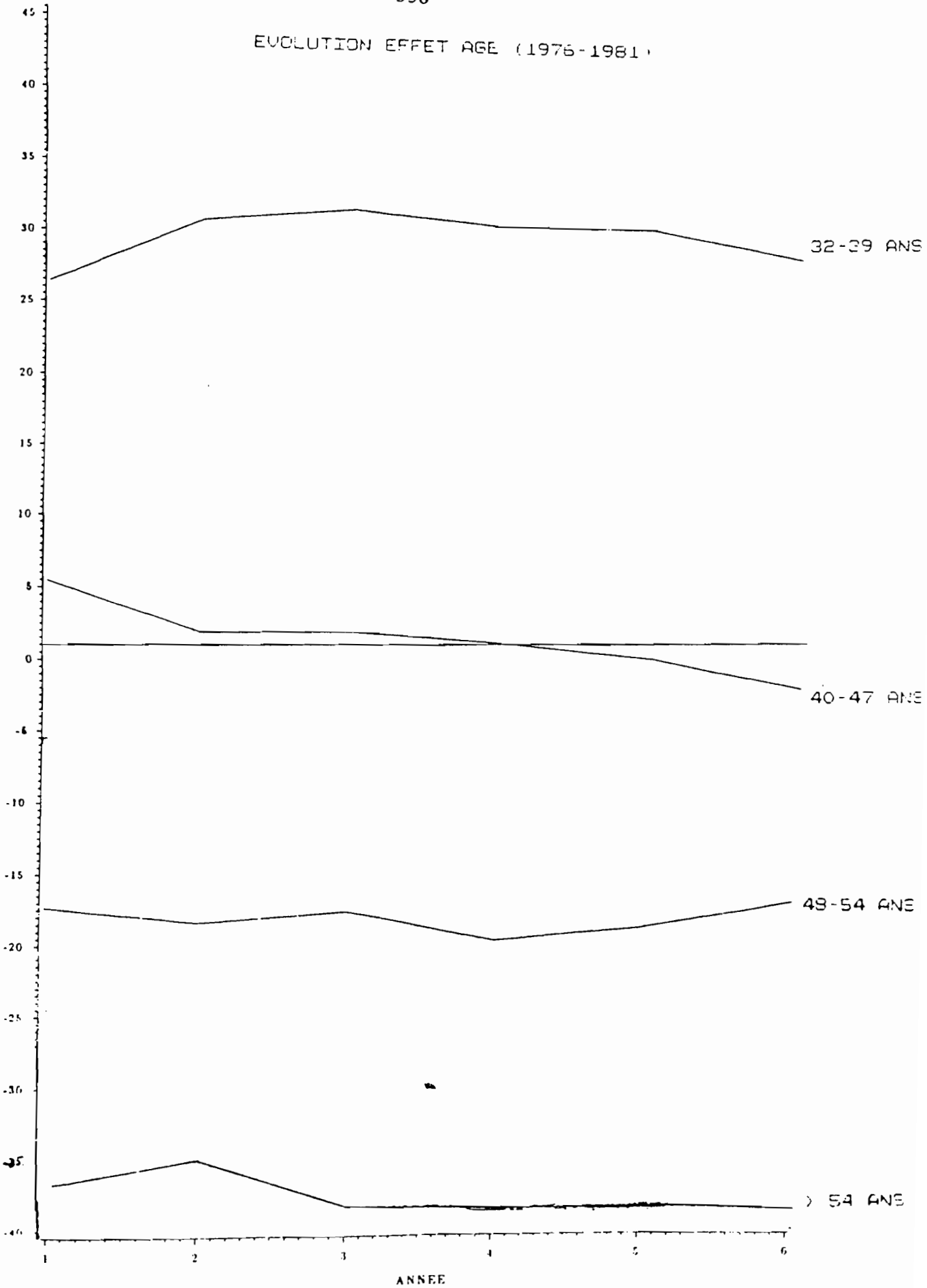
EVOLUTION FEMINISATION RELATIVE DES PROFESSIONS (1976-1991)



EVOLUTION EFFET REGION (1976-1981)

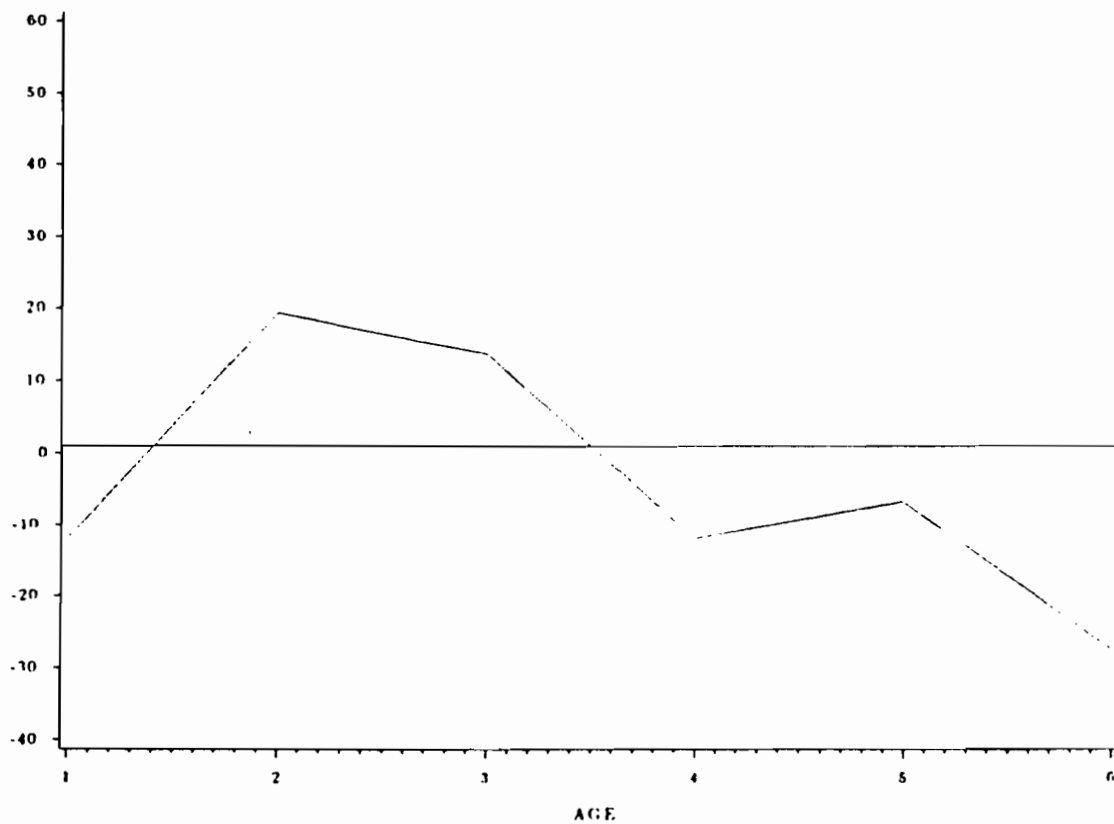


EVOLUTION EFFET AGE (1976-1981)



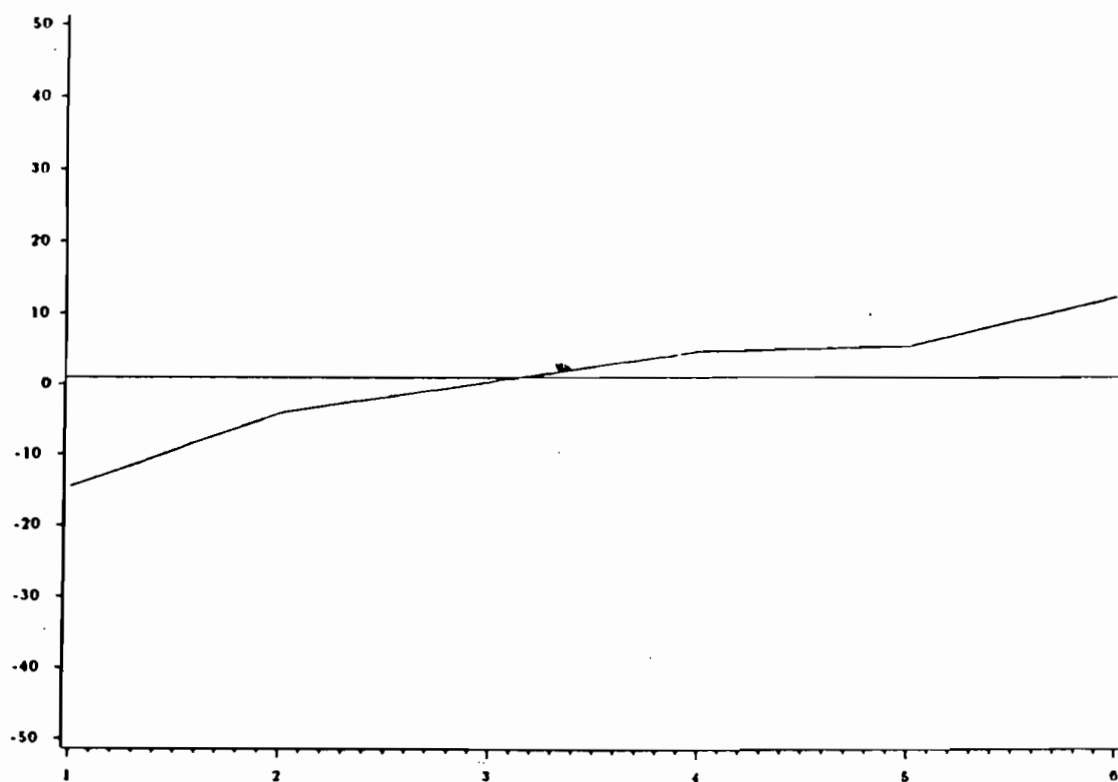
CHAMPAGNE - ARDENNE

PROFIL REG. AGE

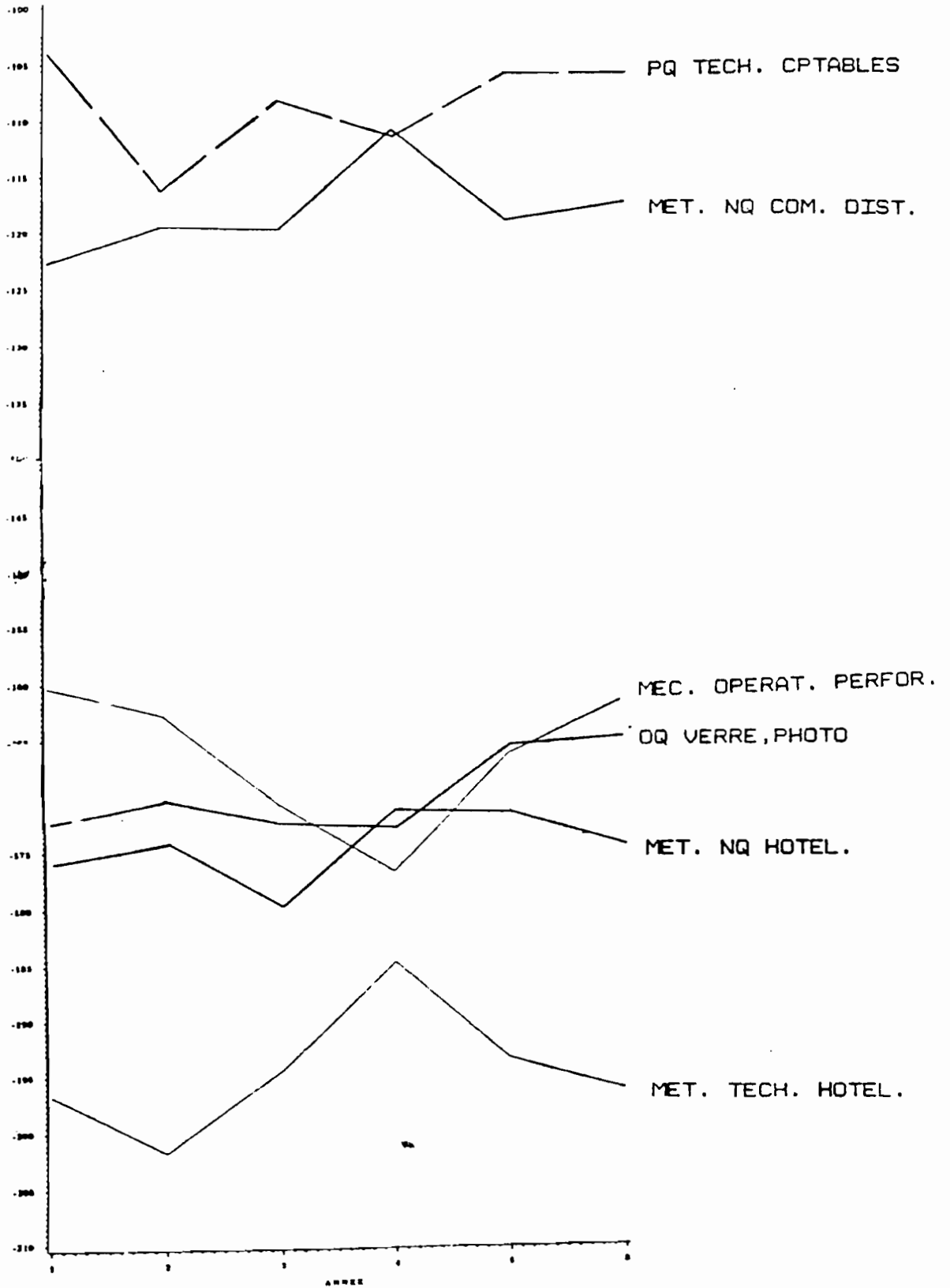


ILE-DE-FRANCE

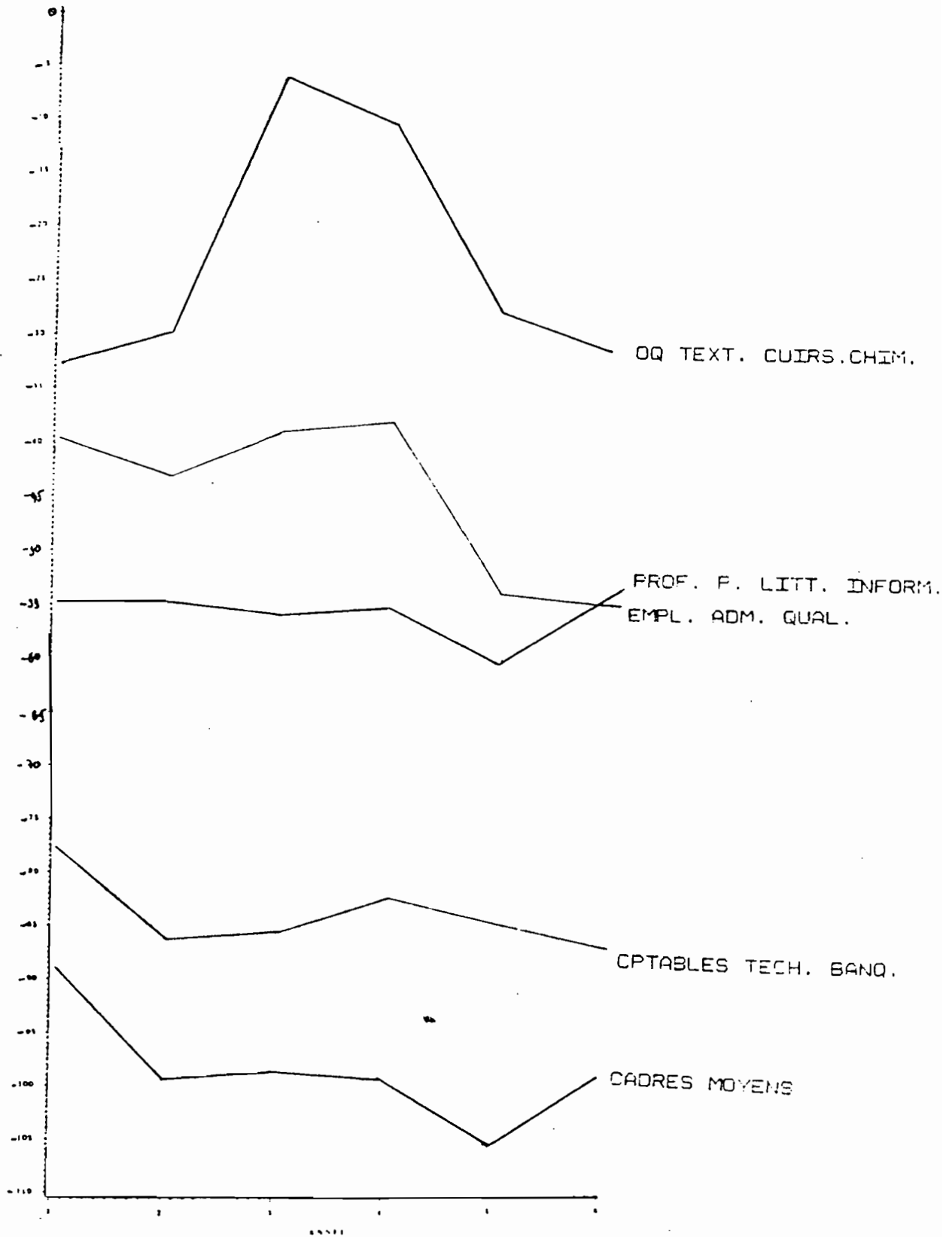
PROFIL REG. AGE



EVOLUTION EFFET PROFESSION (1976-1981)



EVOLUTION FEMINISATION RELATIVE DES PROFESS. (1976-1981)



| S | A | P | R | | | | | TOTAL |
|----|-------|-------|----|----|----|----|-----|-------|
| | | | R1 | R2 | R3 | R4 | R5 | |
| S1 | A1 | P1 | 5 | 5 | 1 | 1 | 1 | 15 |
| | | P2 | 7 | 28 | 4 | 13 | 10 | 64 |
| | | P3 | 2 | 2 | 0 | 1 | 1 | 7 |
| | | P4 | 1 | 1 | 0 | 0 | 0 | 4 |
| | | P5 | 12 | 24 | 4 | 10 | 7 | 58 |
| | | TOTAL | 28 | 61 | 11 | 27 | 21 | 150 |
| | A2 | P1 | 14 | 14 | 3 | 5 | 6 | 44 |
| | | P2 | 8 | 21 | 3 | 9 | 7 | 50 |
| | | P3 | 2 | 4 | 1 | 1 | 2 | 12 |
| | | P4 | 3 | 3 | 0 | 1 | 1 | 12 |
| | | P5 | 19 | 33 | 7 | 13 | 11 | 64 |
| | | TOTAL | 46 | 77 | 17 | 32 | 28 | 204 |
| | A3 | P1 | 12 | 11 | 4 | 4 | 5 | 38 |
| | | P2 | 8 | 16 | 3 | 6 | 4 | 39 |
| | | P3 | 2 | 3 | 1 | 1 | 1 | 10 |
| | | P4 | 3 | 3 | 1 | 1 | 1 | 11 |
| | | P5 | 16 | 26 | 6 | 10 | 10 | 68 |
| | | TOTAL | 43 | 60 | 16 | 24 | 23 | 168 |
| | A4 | P1 | 12 | 12 | 3 | 4 | 4 | 38 |
| | | P2 | 6 | 17 | 2 | 6 | 6 | 41 |
| | | P3 | 2 | 3 | 1 | 1 | 2 | 11 |
| | | P4 | 2 | 2 | 1 | 1 | 1 | 10 |
| | | P5 | 16 | 27 | 5 | 10 | 10 | 70 |
| | | TOTAL | 42 | 64 | 15 | 23 | 25 | 172 |
| A5 | P1 | 10 | 10 | 3 | 4 | 3 | 32 | |
| | P2 | 4 | 14 | 2 | 5 | 4 | 32 | |
| | P3 | 1 | 3 | 1 | 1 | 1 | 10 | |
| | P4 | 2 | 2 | 1 | 0 | 1 | 8 | |
| | P5 | 12 | 19 | 5 | 8 | 8 | 54 | |
| | TOTAL | 31 | 50 | 14 | 20 | 20 | 137 | |
| A6 | P1 | 9 | 7 | 1 | 3 | 3 | 26 | |
| | P2 | 3 | 9 | 2 | 4 | 3 | 23 | |
| | P3 | 2 | 3 | 1 | 2 | 2 | 11 | |
| | P4 | 2 | 1 | 1 | 1 | 1 | 7 | |
| | P5 | 9 | 13 | 5 | 7 | 6 | 41 | |
| | TOTAL | 27 | 35 | 13 | 18 | 16 | 111 | |
| S2 | A1 | P1 | 4 | 3 | 0 | 2 | 1 | 12 |
| | | P2 | 3 | 18 | 1 | 9 | 4 | 36 |
| | | P3 | 6 | 11 | 2 | 5 | 5 | 32 |
| | | P4 | 3 | 4 | 0 | 1 | 1 | 12 |
| | | P5 | 4 | 7 | 1 | 3 | 2 | 18 |
| | | TOTAL | 23 | 44 | 6 | 22 | 15 | 112 |
| | A2 | P1 | 8 | 5 | 0 | 2 | 2 | 20 |
| | | P2 | 3 | 11 | 0 | 5 | 3 | 25 |
| | | P3 | 6 | 10 | 3 | 4 | 5 | 30 |
| | | P4 | 6 | 5 | 1 | 2 | 2 | 17 |
| | | P5 | 4 | 6 | 1 | 3 | 2 | 18 |
| | | TOTAL | 29 | 38 | 8 | 17 | 16 | 112 |
| | A3 | P1 | 6 | 3 | 1 | 1 | 1 | 13 |
| | | P2 | 2 | 6 | 0 | 2 | 2 | 14 |
| | | P3 | 5 | 8 | 3 | 3 | 4 | 24 |
| | | P4 | 5 | 3 | 1 | 1 | 2 | 14 |
| | | P5 | 4 | 4 | 1 | 2 | 1 | 13 |
| | | TOTAL | 24 | 24 | 7 | 11 | 11 | 80 |
| | A4 | P1 | 6 | 2 | 0 | 1 | 1 | 13 |
| | | P2 | 4 | 6 | 0 | 3 | 1 | 19 |
| | | P3 | 5 | 7 | 3 | 4 | 4 | 25 |
| | | P4 | 4 | 2 | 1 | 1 | 1 | 12 |
| | | P5 | 5 | 3 | 0 | 2 | 1 | 13 |
| | | TOTAL | 27 | 25 | 7 | 13 | 11 | 84 |
| A5 | P1 | 4 | 2 | 0 | 1 | 1 | 9 | |
| | P2 | 3 | 5 | 0 | 3 | 2 | 14 | |
| | P3 | 4 | 6 | 2 | 3 | 3 | 20 | |
| | P4 | 3 | 2 | 1 | 1 | 1 | 10 | |
| | P5 | 3 | 3 | 0 | 1 | 0 | 11 | |
| | TOTAL | 19 | 20 | 6 | 11 | 9 | 67 | |
| A6 | P1 | 2 | 1 | 0 | 0 | 0 | 6 | |
| | P2 | 2 | 3 | 0 | 2 | 1 | 11 | |
| | P3 | 4 | 7 | 2 | 4 | 3 | 22 | |
| | P4 | 2 | 1 | 1 | 0 | 1 | 7 | |
| | P5 | 4 | 2 | 0 | 1 | 1 | 10 | |
| | TOTAL | 16 | 17 | 6 | 9 | 8 | 58 | |

ANNEXE A.7

L'INFORMATION DISCRIMINANTE EST DEFINIE TEL QUE

$$D(i_1, i_2) = \sum_{\omega} \left[\frac{t(i_1, \omega)}{t(i_1, +)} - \frac{t(i_2, \omega)}{t(i_2, +)} \right] \text{Log} \frac{t(i_1, \omega)}{t(i_2, +)}$$

LA MATRICE DES PERTES DE DEVIANCE EST DEFINIE PAR :

$$\begin{aligned} PD(i_1, i_2) = & \frac{1}{t(+, +)} \sum \left[\frac{t(i_1, \omega)}{t(i_1, +)} \text{Log} \frac{t(i_1, \omega)}{t(i_1, +)} * \frac{t(i_1 + i_2, \omega)}{t(i_1 + i_2, \omega)} \right. \\ & \left. + \frac{t(i_2, \omega)}{t(i_2, +)} \text{Log} \frac{t(i_2, \omega)}{t(i_2, +)} * \frac{t(i_1 + i_2, \omega)}{t(i_1 + i_2, \omega)} \right] \end{aligned}$$

ANNEXE A8

RUBRIQUES DETAILLEES ET REGROUPEMENTS DES DIFFERENTS METIERS

GROUPE DES METIERS PRIMAIRES

- 11 Agriculteurs et bûcherons
- 12 Marins et pêcheurs

GROUPE DES METIERS SECONDAIRES. NIVEAU I.II (S.I.II)

- 20 Ingénieurs, architectes, scientifiques

GROUPE DES METIERS SECONDAIRES. NIVEAU III.IV (S.III.IV)

- 31 Techniciens et petits entrepreneurs
- 32 Dessinateurs industriels et du bâtiment

GROUPE DES METIERS SECONDAIRES. NIVEAU V

- 4A Agents de maîtrise
- 4B Ouvriers qualifiés et artisans - des mines et carrières
- 4C - de la maçonnerie et des travaux connexes
- 4D - de la couverture , plomberie, sanitaire
- 4E - de la peinture et décoration
- 4F - de la production et première transformation des métaux
- 4G - de la forge, soudure, chaudronnerie charpente en fer
- 4H - de la mécanique
- 4I - de l'électricité
- 4J - de la radio et de l'électronique
- 4K - du verre et de la céramique
- 4L - de la photographie et de l'impression
- 4M - du papier et du carton
- 4N - de la chimie
- 4O - de la boulangerie-pâtisserie
- 4P - de la boucherie-charcuterie
- 4Q - des autres spécialités de l'alimentation
- 4R - des textiles
- 4S - du travail des étoffes
- 4T - des cuirs et peaux
- 4U - du bois et des métiers connexes
- 4V - des conducteurs d'appareils de levage et engins de terrassement
- 4W - de la manutention et magasiniers
- 4X - divers

3Y Conducteurs d'engins de transport

GROUPE DES METIERS SECONDAIRES. NIVEAU VI (S.VI)

50 Ouvriers spécialisés et manoeuvres

GROUPE DES METIERS TERTIAIRES. NIVEAU I.II (T.I.II)

61 Cadres administratifs supérieurs et juristes

62 Médecins et assimilés

63 Professeurs

64 Professeurs littéraires et de l'information

65 Officiers (armée, police, douane, pompiers)

GROUPE DES METIERS TERTIAIRES. NIVEAU III.IV (T.III.IV)

71 Cadres moyens administratifs et juridiques

72 Comptables et techniciens des banques

73 Analystes, programmeurs

74 Métiers techniques de la distribution

75 Métiers techniques de l'hôtellerie

76 Personnels diplômés des services paramédicaux et sociaux

77 Instituteurs et assimilés

78 Artistes

79 Personnel des cultes

GROUPE DES METIERS TERTIAIRES. NIVEAU V (T.V)

81 Employés administratifs qualifiés

82 Personnel de secrétariat

83 Personnel qualifié des techniques comptables

84 Mécanographes, opérateurs, perforateurs

85 Métiers qualifiés de la vente

86 Métiers qualifiés de l'hôtellerie

87 Métiers de la santé (niveau CAP)

88 Métiers des soins personnels

89 Armée, police, douane, pompiers

GROUPE DES METIERS TERTIAIRES. NIVEAU VI (T.VI)

91 Employés non qualifiés

92 Métiers non qualifiés du commerce et de la distribution

93 Métiers non qualifiés de l'hôtellerie

94 Personnel de service

95 Personnel de surveillance

ANNEXE A9

**ÉVOLUTION DE L'EMPLOI TOTAL
AU COURS DES PÉRIODES 1968-1975 ET 1975-1981**

| RÉGION | 1968-1975 | | 1975-1981 | | Différence
des taux
de croissance
moyens |
|-----------------------------------|-----------------------|------|-----------------------|------|---|
| | Taux
de croissance | Rang | Taux
de croissance | Rang | |
| Île-de-France..... | 1,23 | 3 | 0,14 | 15 | - 1,09 |
| Champagne - Ardenne..... | 0,74 | 9 | - 0,14 | 20 | - 0,88 |
| Picardie..... | 1,14 | 4 | 0,17 | 13 | - 0,97 |
| Haute-Normandie..... | 1,00 | 8 | 0,02 | 18 | - 0,98 |
| Centre..... | 1,11 | 5 | 0,63 | 4 | - 0,48 |
| Basse-Normandie..... | 0,23 | 16 | 0,61 | 6 | 0,38 |
| Bourgogne..... | 0,58 | 12 | 0,18 | 14 | - 0,40 |
| Nord - Pas-de-Calais..... | 0,34 | 14 | - 0,46 | 22 | - 0,80 |
| Lorraine..... | 0,56 | 12 | - 0,42 | 21 | - 0,98 |
| Alsace..... | 0,68 | 10 | 0,63 | 4 | - 0,05 |
| Franche-Comté..... | 1,07 | 6 | 0,03 | 17 | - 1,04 |
| Pays de la Loire..... | 0,60 | 11 | 0,44 | 9 | - 0,16 |
| Bretagne..... | - 0,08 | 20 | 0,13 | 16 | 0,21 |
| Poitou - Charentes..... | 0,03 | 19 | 0,72 | 2 | 0,69 |
| Aquitaine..... | 0,07 | 18 | 0,41 | 10 | 0,34 |
| Midi - Pyrénées..... | 0,13 | 17 | 0,49 | 8 | 0,36 |
| Limousin..... | - 0,73 | 22 | 0,21 | 12 | 0,94 |
| Rhône - Alpes..... | 1,06 | 7 | - 0,03 | 19 | - 1,09 |
| Auvergne..... | - 0,21 | 21 | 0,56 | 7 | 0,77 |
| Languedoc - Roussillon..... | 0,28 | 15 | 1,32 | 1 | 1,04 |
| Provance - Alpes - Côte d'Azur... | 1,63 | 2 | 0,64 | 3 | - 0,99 |
| Corse..... | 3,23 | 1 | 0,28 | 11 | - 2,95 |
| France..... | 0,74 | - | 0,23 | - | - 0,51 |

SOURCE : "Disparités et diversité des régions françaises à la veille du 9ème Plan", Archives et Documents.

FÉMINISATION DE L'EMPLOI SALARIÉ DE 1968 A 1980

| RÉGION | Part des femmes dans les secteurs industriel et tertiaire (1) | | | | Croissance des deux de féminité (1) | | | | | | Part des femmes dans l'emploi salarié féminin | | | | |
|-------------------------------------|---|------|-----------|------|-------------------------------------|-----------|-----------------|-----------|----------------------|-------------------|---|------|------|------|-----|
| | Industrie - Bâtiment | | Tertiaire | | Industrie | | Tertiaire | | Valeur en 31-12 1968 | Indices de départ | Croissance de la part des effectifs féminins | | | | |
| | 1968 | 1974 | 1968 | 1974 | 1968-1974 | 1974-1980 | Total 1968-1980 | 1968-1974 | | | 1974-1980 | % | % | | |
| Ile-de-France..... | 27,6 | 27,2 | 26,5 | 51,8 | 50,6 | 49,5 | 0,4 | 0,7 | 1,1 | 1,2 | 1,1 | 2,3 | 1,06 | 2,3 | 4,4 |
| Champagne - Ardenne..... | 27,8 | 27,9 | 26,3 | 49,9 | 46,2 | 42,5 | -0,1 | 1,6 | 1,5 | 3,7 | 3,7 | 68,0 | 0,91 | 2,3 | 7,4 |
| Picardie..... | 26,2 | 26,6 | 25,3 | 51,6 | 47,9 | 44,3 | -0,4 | 1,3 | 0,9 | 3,7 | 3,6 | 67,6 | 0,90 | 2,4 | 7,7 |
| Haute-Normandie..... | 25,0 | 25,9 | 23,8 | 51,5 | 48,9 | 46,0 | -0,9 | 2,1 | 1,2 | 2,6 | 2,9 | 5,5 | 0,95 | -0,6 | 6,3 |
| Centre..... | 29,8 | 29,4 | 26,3 | 51,3 | 49,4 | 46,5 | +0,6 | 3,1 | 3,7 | 1,9 | 2,9 | 4,8 | 0,90 | - | 4,5 |
| Basse-Normandie..... | 25,2 | 25,5 | 20,9 | 52,3 | 51,0 | 49,6 | -0,3 | 4,6 | 4,3 | 1,3 | 1,4 | 2,7 | 0,97 | 1,7 | 4,7 |
| Bourgogne..... | 25,7 | 25,7 | 22,2 | 50,4 | 48,1 | 45,3 | - | 3,5 | 3,5 | 2,3 | 2,8 | 5,1 | 0,96 | -0,5 | 5,2 |
| Nord - Pas-de-Calais..... | 21,3 | 22,0 | 21,0 | 48,8 | 46,9 | 44,6 | -0,7 | 1,0 | 0,3 | 1,9 | 2,3 | 4,2 | 0,95 | 4,2 | 8,0 |
| Lorraine..... | 19,8 | 18,9 | 16,7 | 47,9 | 45,6 | 42,6 | +0,9 | 2,2 | 3,1 | 2,3 | 3,0 | 5,3 | 0,98 | 1,6 | 5,5 |
| Alsace..... | 24,8 | 24,8 | 22,9 | 48,5 | 45,9 | 42,7 | - | 1,9 | 1,9 | 2,6 | 3,2 | 5,8 | 0,94 | 0,9 | 4,9 |
| Franche-Comté..... | 26,8 | 27,0 | 27,4 | 51,9 | 48,9 | 45,7 | -0,2 | -0,4 | -0,6 | 2,0 | 3,2 | 5,2 | 0,94 | 4,3 | 7,5 |
| Pays de la Loire..... | 27,7 | 27,1 | 23,1 | 51,6 | 50,7 | 48,9 | +0,6 | 4,0 | 4,6 | 0,9 | 1,8 | 2,7 | 0,97 | -1,5 | 3,4 |
| Bretagne..... | 23,3 | 24,1 | 19,5 | 46,4 | 46,2 | 43,5 | -0,8 | 4,6 | 3,8 | 2,2 | 2,7 | 4,9 | 1,04 | -0,9 | 4,9 |
| Poitou - Charentes..... | 24,6 | 24,5 | 20,0 | 50,0 | 47,9 | 45,0 | +0,1 | 4,5 | 4,6 | 2,1 | 2,9 | 5,0 | 0,99 | -2,1 | 4,2 |
| Aquitaine..... | 22,1 | 22,8 | 21,7 | 50,6 | 47,9 | 45,5 | -0,7 | 1,1 | 0,4 | 2,7 | 2,4 | 5,1 | 1,05 | 2,4 | 5,0 |
| Midi - Pyrénées..... | 23,4 | 22,6 | 19,5 | 50,1 | 48,2 | 45,4 | +0,8 | 3,1 | 3,9 | 1,9 | 2,8 | 4,7 | 1,04 | 1,3 | 3,7 |
| Limousin..... | 26,9 | 27,2 | 26,6 | 51,2 | 47,8 | 44,2 | -0,3 | 0,6 | 0,3 | 3,4 | 3,6 | 7,0 | 0,97 | 2,1 | 5,1 |
| Rhône - Alpes..... | 26,1 | 26,4 | 26,0 | 50,6 | 48,5 | 46,3 | -0,3 | 0,4 | 0,1 | 2,1 | 2,2 | 4,3 | 0,93 | 3,9 | 6,8 |
| Auvergne..... | 24,8 | 23,5 | 21,4 | 53,3 | 50,1 | 47,4 | +1,3 | 2,1 | 3,4 | 3,2 | 2,9 | 5,1 | 0,97 | 1,2 | 4,9 |
| Languedoc - Roussillon..... | 18,1 | 17,9 | 15,5 | 46,4 | 44,1 | 41,6 | +0,2 | 2,4 | 2,6 | 2,3 | 2,5 | 4,8 | 1,12 | 1,8 | 3,2 |
| Provence - Alpes - Côte d'Azur..... | 15,3 | 14,6 | 15,3 | 45,7 | 43,8 | 41,2 | +0,7 | -0,7 | - | 1,9 | 2,6 | 4,5 | 1,15 | 3,1 | 2,7 |
| Corse..... | 24,8 | 24,7 | 23,1 | 50,3 | 48,3 | 46,1 | 1,6 | +0,1 | 1,7 | 2,2 | 2,0 | 4,2 | 1,00 | 1,7 | 5,1 |
| France..... | 24,8 | 24,7 | 23,1 | 50,3 | 48,3 | 46,1 | 1,6 | +0,1 | 1,7 | 2,2 | 2,0 | 4,2 | 1,00 | 1,7 | 5,1 |

(1) En % de l'emploi salarié féminin.

SOURCE : "Disparités et diversité des régions françaises à la veille du 9ème Plan", Archives et Documents.

BIBLIOGRAPHIE

1. ABRAHAM-FROIS G. (1986) : Economie politique, 3^{ème} Edit., Economica.
2. ACZEL J., DAROCZ Z.(1975) : On measure of information and characterisations Academic Press, N.Y.
3. ADAMS J.D.(1985) : Permanent differences unemployment and permanent wage differentials, **Quarterly Journal of Economics**, pp. 29-55.
4. AGLIETTA M.(1982) : « Panorama sur les théories de l'emploi » in *Emploi et Chômage*, Presses de la Fondation Nationale des Sciences Politiques.
5. AGRESTI A.(1983) : Survey of strategies for modelling cross-classification having ordinal variables, **JASA**, Mars , vol. 381.
6. AMAT F., GEHIN J.P.(1985) : Insertion professionnelle des jeunes et mobilité des actifs dans un contexte de crise de l'emploi, Communication aux journées d'étude sur les structures du marché de travail et les politiques de l'emploi, 3-4, Octobre, Paris.
7. ANDERSEN A.H(1974) : Multidimensional contingency tables, **Scand. Journal Stat.** Vol. 1, pp.115-127.
8. ANDERSEN E.B.(1980) : Discrete statistical models with social science applications, North-Holland.
9. ANDERSON J.A.(1984) : Regression and ordered categorical variables, **JRSS**, Vol. 46 n°1, pp. 1-30.
10. ANDRICH D.(1979) : A model for contingency tables having an ordered response classification, **Biometrics**, Vol. 35 pp. 403-415.
11. ARMANO C., TERRIER C.(1987) : « Etudier le local avec l'INSEE », *Données sociales*, pp. 348-352.
12. AYDALOT PH.(1976) : Dynamique spatiale et développement inégal, *Economica*
13. AYDALOT PH.(1978) : Mobilités des activités et de l'emploi, **RERU**, n° 3
14. AYDALOT PH.(1983) : Le rôle du travail dans les nouvelles stratégies de localisation, **RERU**, n° 2, pp. 175-189.
15. AYDALOT PH.(1983) : « Division spatiale du travail » in *Espace et localisation*, *Economica*.
16. AYDALOT PH.(1984) : « A la recherche des nouveaux dynamismes spatiaux » in *Crise et Espace*, *Economica*, pp. 38-59.

17. AZOULI A.(1981) : « Théorie et pseudo-théorie : le dualisme du marché du travail » in Segmentation de l'emploi ou division du travail ? Critiques de l'économie politique n° 15/16 pp. 3-52.
18. BECKER R.J, NELDER J.A(1978) : The Generalized LINEAR Interactive Modelling (GLIM), Release 3, Oxford NAG.
19. BARTLETT (1935) : Contingency table interaction, **JRSS** n° 2, pp. 248-252.
20. BARNDORFF, NIELSON O.(1978) : Information and exponential families in statistical theory, N.Y., Wiley & Sons.
21. BELLOC B, MARCHAND O.(1986) : L'emploi régional depuis 1967, **Eco & Stat** n°193/194, Nov-Déc, pp. 21-126.
22. BENASSY J. P.(1976) : Théorie du déséquilibre et fondements microéconomiques, **Revue Economiques**.
23. BENHAYOUN G.(1984) : L'emploi et le chômage des jeunes : Analyse urbaine.
24. BENZECRI (1984) : L'analyse des données Tome1 : Taxinomie, Dunod.
25. BHAPKAR V.P, KOCH G.G.(1968) : Hypotheses of no interaction in contingency tables, **Biometrics**, Vol. 24, pp. 517-594.
26. BIRCH M.W.(1965) : The detection of partial association, II : the general case **JRSS B**, n°27, pp. 111-124.
27. BISHOP Y.M, FIENBERG S.E, HOLLAND P.W.(1975) : Multivariate statistical analysis, theory and practice, MIT Press, Cambridge Mass.
28. BLAUG M.(1981) : La pensée économique, 3^{ème} Edit., **Economica**.
29. BONNETT D.G, BENTLE P.M.(1983) : Goodness of fit procedure for the evaluation and selection of log-linear models, **Psychological Bulletin**, Vol. 93, n° 1, pp. 149-166.
30. BOUCHET J.(1984) : « Dynamique spatiale de la population et des activités. Questions pour l'aménagement du territoire » in Crise et Espace, **Economica** pp 60-79.
31. BOUDEVILLE J. R.(1968) : L'espace et les pôles de croissance, PUF.
32. BOUILLAGUET B, GAUVIN A., OUTIN J. L.(1981) : Femmes au travail prospérité et crise, **Economica**.
33. BOUISSOU M. B, LAFFONT J. J, VUONG Q. H.(1984) : Econométrie du déséquilibre sur microéconomiques, **Annales de l'INSEE**, n° 55/56.
34. BOYER R.(1984) : Les transformations du rapport salarial en Europe. Quelques éléments de synthèse 1973 à 1984, **FERE**, Août 1984.

35. BREEN R.(1984) : Fitting non hierarchical and association models using GLIM, **Sociological Methods and research**, n° 13 pp. 77-107.
36. BREEN R.(1985) : Log-Multiplicative models for contingency tables using GLIM, **Newsletter 10**, pp. 14-19.
37. BUTTHER O., MOURIAUX M. F.(1984) : Le développement des services marchands aux entreprises : mouvement d'extériorisation ou croissance autonome ? in **Cahiers du CEE**, n° 27, Vol 2.
38. CAMPAGNA E.(1983) : La différenciation des espaces locaux dans les politiques de main-d'œuvre dans les grandes entreprises. L'exemple d'USINOR-DUNKERQUE, **RERU**, n°1 pp. 8-25.
39. CEE (1983) : Une stratégie locale pour l'emploi, Colloque Ministère de l'emploi, **Bulletin d'Information**, n° 60/61.
40. CENTI C., LAITON C., TEMPIERE E.(1988) : Pour un repérage des zones d'emploi complexes, **CERS**.
41. CHALMERS J.A, BECKHELM T. L(1976) : « Shift and Share the theory of industrial location » in **Regional Studies**, Vol. 10, pp 15-23.
42. CHAMBERS J. M.(1973) : Fitting non linear models: Numerical techniques, **Biometrika**, Vol. 60, p. 1.
43. CLOGG C. C.(1982a) : Using association models in sociological research: some examples, **JAS**, Vol. 88 n° 1, pp. 114-134.
44. CLOGG C. C.(1982 b) : Some models for the analysis of association in multiway cross-classifications having ordered categories, **JASA**, Vol. 74, pp. 537-552.
45. CLOGG C. C, GOODMAN L. A.(1984) : Latent structure analysis of a set of multidimensional contingency tables, **JASA**, Vol. 79, n° 388.
46. COEFFIC N.(1982) : D'octobre 1980 à octobre 1981 l'aggravation du chômage s'est poursuivie, **Eco & Stat**, n° 145 pp 29-40.
47. COURBIS R, PRAGER J. C.(1973) : Analyse régionale et planification nationale le modèle REGINA, **Collections de l'INSEE**, Série R, n° 12.
48. CSISZARD I. KORNER.(1981) : Information theory: coding theorems for discrete memoryless systems, **academic Press N. Y.**
49. CUNNINGHAM A.D.(1984a) : The concept of distance between structures and surprises values, **Eurostat**.
50. CUNNINGHAM A. D.(1984b) : An analysis of the operation of comparison with application to index numbers, **Eurostat**.

51. DALE W. O and GRIZZLE J. E.(1972) : Analysis of contingency tables having ordered response categories, **JAS A**, Vol. 67, n° 337.
52. DARROCH H. N, RATTCLIF D.(1972) : Generalized iterative scaling for log-linear models, **The Annals of Mathematical Statistics**, Vol. 43, n° 5, pp. 1470-1480.
53. DAUDIN J. J.(1980) : Partial association measures and an application to qualitative regression. **Biometrika**, Vol. 67, n° 3, pp. 581-590.
54. DAUDIN J. J et TRECOURT P.(1980) : Analyse factorielle des correspondances et modèles log-linéaires ; comparaison des deux méthodes sur un exemple. **Revue de Statistique Appliquée**, n°1, pp. 5-24.
55. DESTEFANIS M., VASSEUR A. M.(1974) : Le fonctionnement d'un marché du travail local : le bassin de main-d'œuvre d'Annecy, **Cahier CEE**, n° 5, PUF
56. DESTEFANIS M., FOUCHER L, GOUTERMANOFF(1977) : Le fonctionnement d'un marché du travail local : le bassin de main-d'œuvre de Compiègne, **cahiers CEE**, n° 13, PUF
57. DIJK J.V, FOLMER H.(1986) : The consequences of interregional labor migration for the regional labor market : theory methodology and experience, **The R.E.S Review of Economics and Statistics**, pp. 74-83.
58. DIRECTION DE LA PREVISION(1983) : Vers une précarisation accrue des jeunes sur le marché du travail, DF, 89. BD4. 299/90.
59. DRUFMAN B., MOUY PH.(1985) : Insertion des jeunes : efficacité et limites des politiques de l'emploi, Communication aux journées d'étude sur les structures du marché du travail et les politiques d'emploi 3-4 Oct., PARIS.
60. DUNCAN D.O.(1979) : How destination depends on origin in the occupational mobility table, **A.J.S**, Vol. 84, n° 4.
61. DURR J.M et ABOULIN M.(1985) : « L'âge des français : contrastes régionaux et opposition ville-campagne », **Eco & Stat**, n° 173 pp. 25-32.
62. DUTAILLY J. C.(1981) : La crise du système productif, **Eco & Stat**, n° 138, Nov. 1981 pp. 3-20.
63. EDWARDS J. A.(1976) : Industrial structure and regional change : a shift share analysis of the British Columbia Economy 1961-1970, **Regional Studies**, Vol. 10, pp. 307-317, Pergamon Press, 1976.
64. EHRENBERG R.G, SMITH R. S.(1982) : Modern labor economics, Scott Foresman and Company, Glenview, Illinois
65. EMERSON P. L.(1968) : Numerical construction of orthogonal polynomials from a general recurrence formula, **Biometrics**, pp. 645-700.

66. EUROSTAT(1984) : Numéro spécial. Développements récents dans l'analyse des grands ensembles de données.
67. EYMARD-DUVERNAY F.(1975) : Analyse des liens emploi-chômage, **Eco & Stat**, n°69 Juillet-Août.
68. EYMARD-DUVERNAY F.(1979) : Combien d'actifs d'ici l'an 2000, **Eco & Stat**, Septembre 1979.
69. EYMARD-DUVERNAY F.(1981) : Les secteurs de l'industrie et leurs ouvriers, **Eco & Stat**, n° 138.
70. FAVEREAU O.(1985) : Evolution récente des modèles et des représentations théoriques du fonctionnement du marché du travail, Colloque structures du marché du travail et politiques de l'emploi 3-4, Octobre, PARIS.
71. FAVEREAU O., SOLLOGOUB M.(1986) : «Structure sectorielle de l'emploi et marché du travail»in l'enquête sur les forces de travail comme instrument de la politique de l'emploi, CEE, AVIGNONO.
72. FAY R. E, GOODMAN L.A(1975) : ECTA Program : Description for users, Univ. of Chicago Dept. Of Statistics, Chicago Ill.
73. FIENBERG S. E.(1979) : The use of chi-square statistics for categorical data problem, **JRSS B**, n° 1, pp. 54-64.
74. FISHER R. A.(1934) : Probability likelihood on quantity of information in logic of uncertain inference. Proceeding of the R.S.S of London, A. 146, pp. 1-8
75. FISHER W. D.(1958) : On grouping for maximum homogeneity, **JASA**, 53, pp. 789-798
76. FITOUSSI J.P.(1979) : Emploi, structure et régulation, **Revue d'Economie Politique**, n°1.
77. GAMBIER D.(1978) : « L'intégration de l'espace dans le fonctionnement du marché du travail : la notion du marché du travail : la notion du marché local » in Marchés locaux du Travail, Documentation Française, pp. 39-74.
78. GAMBIER D, VERNIERES M.(1982) : Le marché du travail, *Economica* 2^{ème} Ed.
79. GARNIER O.(1985) : Marchés internes et marché externe du travail : leur rôle dans la formation des salaires, Uni. De recherche INSEE
80. GERARD G. , ROLIN J. M.(1979) : Analyse des données discrètes, Comité de Stat Univ. Catholique de Louvain.
81. GERME J. F, MICHON F.(1979) : Le fonctionnement des marchés de l'emploi : éléments d'analyse, SET, PARIS.

82. GERME G. F.(1981) : « Instabilité, précarité et transformations de l'emploi » in Segmentation de l'emploi ou division du salariat ? Critiques de l'Economie Politique n° 15/16 pp 3-52.
83. GENSTAT(1977) : A General Statistical Program, Harpenden : Rothamsted Experimental Station.
84. GIRAN J. P.(1979) : Recherche d'information et déséquilibres, *Economica*.
85. GIRAN J.P, GRANIER R.(1983) : Politique de l'emploi, *Economica*.
86. GOKHALE D. V.(1971) : An iterative procedure for analysing log-linear models **Biometrics**, n° 27, pp. 681-687.
87. GOKHALE D.V.(1972) : Analysis of log-linear models.
88. GOKHALE D. V, KULLBACK S.(1978) : The information in contingency tables, Marcel Dekker.
89. GOOD I. J.(1963) : Maximum entropy for hypotheses formulation, spatially for multidimensional contingency tables, **The Annals of Mathematical Statistics**, 34, pp. 911-934.
90. GOODMAN L.A.(1970) : The multivariate analysis of qualitative data : interactions among multiple classifications, **JASA**, n° 65, pp. 226-255.
91. GOODMAN L.A.(1971) : The analysis of multidimensional contingency tables : Stepwise procedures and direct estimation methods for building models for multiple classification, **Technometric**, n°13, pp. 33-61.
92. GOODMAN L. A.(1971) : Partitioning of chi-square analysis of marginal contingency tables and estimation of expected frequencies in multidimensional tables, **JAS A**, Vol. 66, pp. 339-344.
93. GOODMAN L. A.(1979) : Multiplicative models for the analysis of occupational mobility tables and others kinds of cross-classifications tables, **AJS**, Vol. 84, n° 4.
94. GOODMAN L. A.(1979) : Simple models for analysis of association in cross-classifications having ordered categories, **JAS A**, Vol. 74, n° 367-552.
95. GOODMAN L. A.(1981) : Association models and canonical correlations in the analysis of cross-classifications having ordered categories, **JAS A**, 320-334.
96. GOODMAN L. A.(1981) : Association models and the bivariate normal distribution in the analysis of cross-classifications having ordered categories, **Biometrika**, n°68, pp 347-355.
97. GOODMAN L. A, KRUSKALL W. H.(1963) : Measures of association for cross-classifications, **JASA**, Vol. 58, pp. 300-364.
98. GOURIEROUX C.(1984) : Econométrie des variables qualitatives, *Economica*.

99. GRIZZLE J. S, STARMER C. F, KOCH G. G.(1969) : Analysis of categorical data by linear models, **Biometrics** 25, pp. 489-504.
100. GUIOT B., HERAN O.(1988) : Les jeunes et leur région, **Archives et Documents**, n°203, INSEE.
101. HABER M.(1984): Fitting a general log-linear model, *Applied Stat.*, pp. 358.
102. HABERMAN S.(1974) : Log-linear models for frequency tables with ordered classifications, **Biometrics**, Dec. pp 589.
103. HABERMAN S.(1974) : The Analysis of frequency data, Chicago.
104. HALL E. R.(1970) : Why is unemployment rate so high at full employment ? Brookings papers on economic activity, pp. 369-384.
105. HAUSER R. M, MASSAGLI M. P.(1983) : Some models of agreement and disagreement in repeated measures of occupation, **Demography**, vol. 20 n° 4.
106. HECKMAN J. J, MALURDY T.(1980) : A dynamic model of female labor force supply, **RES**.
107. HELLER J. L.(1985) : Emploi et chômage en 1985, **Eco & Stat.** n° 183, pp. 21-36
108. HERRAND A., MOUGEOT M.(1978) : Migrations alternantes et bassins d'emploi l'exemple de l'Alsace, **RERU**, n° 2, pp. 193-214.
109. HUET M.(1982) : La progression de l'activité féminine est-elle irréversible ? **Eco & Stat**, n° 145.
110. HUET M.(1983) : La concentration des emplois féminins, **Eco & Stat.**, n° 153, Avril 1983.
111. HUET M.(1984) : La gestion de l'emploi féminin et masculin obéit-elle à des logiques différentes ? Colloque : les femmes et le sexisme.
112. HUTCHINSON D. A.(1985) : Ordinal variables regression using the McCullagh (proportional odd) model, **GLIM Newsletter**, n° 9, pp. 9-17.
113. ISRAELS A. Z et alï.(1984) « Méthodes d'analyse multidimensionnelle pour variables discrètes » in *Développements récents dans l'analyse des grands ensembles de données*, Eurostat, n° spec., pp.241.
114. JAE-ON K.(1975) : Multivariate analysis of ordinal variables, **Amer. Jour. Of Stat.**, Vol. 81, n° 2, pp. 261-294.
115. JAYET H.(1985) : Les zones d'emploi et d'analyse locale du marché du travail **Eco & Stat**, n° 182, pp. 37-42.

116. KAUFMAN R. L, SPILERMAN S.(1982) : The age structure occupations and jobs, **Americ. Jour. Of Sociology**, Vol. 87, n° 4, p 827.
117. KERR C.(1954) : The balkanisation of labor markets, Labor Mobility and Economic Opportunity, Essay by E. W. BAKKE & all MIT Press. N. Y.
118. KOENIG H., NERLOVE M., OUDIZ G.(1979) : Modèles log-linéaires pour l'analyse des données qualitatives, **Annales de l'INSEE**, Vol. 36, pp. 31-81.
119. KOENIG H., NERLOVE M., OUDIZ G.(1981) : On the formation of price expectations : an analysis of business test data by log-linear probability models, **European Economic Review**, Vol. 16, p. 103.
120. KULLBACK S.(1978) : Information theory and statistics, Peter Smith.
121. KULLBACK S, KU H. H.(1968) : Interaction in multidimensional contingency tables : an information theoretic approach, Jour. Res. Nat, Bur. Stand. Sect. Ser. B. 72, pp. 159-199.
122. LACHAT D.(1977) : La filiation des services, **Revue des Sociétés** p. 52.
123. LANCASTER K(1971) : Consumer demand. A new approach. N. Y Colum. Uni. Press
124. LANCASTER T. , NICKELL(1980) : The analysis of re-employment probabilities for the unemployed, **JRSSA**, 143 pp. 141-165.
125. LANCRY P. J.(1982) : Théorie de l'information et économie, *Economica*.
126. LEBORGNE D., LIPIETZ A.(1987) : New technologies, new modes of regulation some special implications, International Conference Technology Restructuring and urban/regional development Dubronik.
127. LECLERC R., VALOIS M.(1978) : Fonctionnement des marchés locaux de l'emploi. Analyse du cas de Rouen. Rapport d'études pour l'aménagement de la basse vallée de Seine.
128. LEE S. K.(1977) : On the asymptotic variance of the u-terms in log-linear models of multidimensional contingency tables, **JASA**, Vol. 72, n° 358, pp. 412-419.
129. LEHRER E. L, STOKES H.(1984) : Determinants of female occupational distribution : a log-linear probability analysis, **Rev. Eco. Stat.**, pp. 395-404.
130. LEROY R.(1981) : Le marché du travail : une approche hors paradigmes, **Revue Economique**, Vol. 32, n° 2, pp. 249-270.
131. LIPIETZ A.(1978) : La dimension régionale du développement du tertiaire, **Cahiers du CEPREMAP**.
132. LOINGER G.(1982) : De l'économie à l'économie spatiale. Note de travail.

133. LOLLIVIER S.(1984) : Revenu offert, prétentions salariales et activités des femmes mariées, un modèle d'analyse, **Eco & Stat**, n° 1.
134. LOVERIDGE R., MOK A.(1981) : « Théories de la segmentation du marché du travail » in *Dossiers Marché du Travail Européen, Série Politique Sociale*, n° 42, pp. 11-28.
135. LUNNEBORG C. E, ABBOT R. D.(1983) : Elementary multivariate analysis of behavioral sciences. Application of basic structure. North Holland.
136. MABILE S. JAYET H.(1985) : La redistribution géographique des emploi entre 1975 et 1982, **Eco & Stat** , n° 182, pp. 23-36.
137. McCULLAGH, NELDER J. A.(1983) : Generalized linear models. Chapan & Hall.
138. MACLOUF P., SARBIB J.(1983) : Dualisation spatiale et dualisation social de la force de travail en période de crise. Deux scénario possibles, **RERU**, n° 1 1983, pp. 119-146.
139. MADINIER H.(1984) : Analyse de l'emploi salarié par zones d 'emploi Colloque A. E. A-A. S. R. D. L. F sur la modélisation régionale spatiale.
140. MAGIDSON J. J, SWAN J. H, BERK R. A.(1981) : Estimating non hierarchical arnested log-linear models, **Soc. Meth. And Research**, Vol. 10, pp. 3-49.
141. MALINVAUD E.(1984) : « Présentation » in *Développements récents dans l'analyse de grands ensembles de données* , **Eurostat**, p. 7.
142. MALLET L.(1978) « Marché de la force de travail et localisation » : in *Marchés locaux du travail*, **Documentation Française**, pp.75-97.
143. MALLET L.(1980) : Le marché local du travail, Edition du CNRS.
144. MALLET L, VERNIERES M.(1981) : L'expérience professionnelle dans les modes de gestion de la main-d'œuvre, **Travail et emploi**, n° 9, pp. 65-74.
145. MARCHAND O, REVOIL J. P.(1981) : Emploi et Chômage, bilan fin 1980, **Eco Stat**. n°130.
146. MARCHAND O, MARC N.(1984) : La population active de 1975 à 1982 : les facteurs d'une croissance, **Eco & Stat**, n° 171-172, pp. 5-24.
147. MARCHAND O, MARTIN L. G.(1986) : L'emploi en 1985 : les signes d'une amélioration, **Eco & Stat**, n° 187, pp. 47-58.
148. MARSTON S.(1985) : Two views of the geographic distribution unemployment, *Quarterly Journal of Economics*, pp. 57-79.
149. MEDOFF J.(1983) : U. S labor markets : imbalances, wage growth an productivity in 70's. *Brookings Papers on Economic Activity*, n° 1.

150. MERIAUX M.(1982) : « Point de vue sur les recherches françaises en économie du travail » in *Emploi et Chômage*, Presse de la Fondation Nationale des Sciences Politiques, pp. 120-140.
151. MICHAU M.(1978) : « Marché du travail local et bassins d'emploi : Etat de études et évolution de la recherche » in *Marchés locaux au travail*, Documentation Française pp. 99-118.
152. MONNOYER M. C.(1982) : Détermination des zones vulnérables dans un espace industriel régional : application à la Rhône-Alpes, **RERU**, n° 3, pp. 363-386.
153. NEGRE M.(1981) : Modèles macroéconomiques interrégionaux et « seuils de complexité », **RERU**, n° 2, pp. 213-227.
154. NELDER J. A.(1974) : Log-linear models for contingency tables, a generalization of classical least squares, **Applied statistics**, Vol. 23, n° 3, pp. 323-329.
155. NELDER J. A.(1984) : « Le rôle des modèles dans les statistiques officielles » in développements récents dans l'analyse de grands ensembles de données , **Eurostat**, Numéro Spécial.
156. NELDER J. A, WEDDERBURN R. W.(1972) : Generalized linear models, **JRSS A**, n° 135 pp. 370-384.
157. NERLOVE N., PRESS S. J.(1975) : Multivariate log-linear probability models for the analysis of qualitative data. **Discussion paper**, n° 1, Center of Statistics and Probability North Western.
158. NOHARA H.(1983) : Unité et dualité du marché du travail industriel, **LEST**.
159. OTTENWALTER B, VUONG O.(1981) : Modèles conditionnels log-lineaires de probabilité et systèmes récurrents, **ENSAE**.
160. OTTENWALTER B., OUDIZ G.(1982) : Anticipations et décisions des entreprises : une étude économique des réponses aux enquêtes de conjoncture de l'INSEE. Rapport Scientifique ASP 6087 CNRS, NSF, ERA 199.
161. PERRIN J. C.(1983) : « Economie spatiale et méso-analyse », in *Espace et Localisation*, *Economica*, pp. 201-299.
162. PIORE M.(1982) : « Dualism in labor market », in *Emploi et Chômage*, Presse de la Fondation Nationale des Sciences Politiques, pp. 26-48.
163. PLACKETT R. L.(1962) : A note on interactions in contingency tables, **Jour. Of the Royal Stat. Soc. B**. Vol. 24, n° 1, pp. 162-166.
164. PLANQUE B.(1983) : Communication et travail-l'espace dualisé jusqu'où ? **RERU**, n°1, pp. 97-117.

165. PLANQUE B.(1984) : «Technologies nouvelles et réorganisation spatiale » in Crise et Espace, *Economica*, pp. 98-121.
166. PLANQUE B., LAZZERI R.(1980) : Disjonction fonctionnelle et développement local, **RERU**, n° 1, pp. 51-71.
167. POTTIER C.(1984) : « Facteurs de rééquilibrage spatial de l'emploi industriel : les régions françaises face à la crise » in Crise et Espace, *Economica* pp. 122-139.
168. POULAIN M.(1981) : Contribution à l'analyse spatiale d'une matrice de migration interne, *Recherches démographiques, Cahiers n° 3 Univ. Cath. Louvain*.
169. RAULT D.(1984) : Secteurs d'activité : l'évolution des structures de la main-d'œuvre, **Eco & Stat.**, N° 171/172.
170. ROBINSON D. , VINCENS J.(1974) : Etude des comportements sur le marché du travail, **OCDE**.
171. ROSE J.(1984) : En quête d'emploi, *Economica*.
172. SALLEZ A.(1983) : Division spatiale du travail, développement régional polarisé et théorie de la localisation, **RERU**, n° 1, pp. 70-95
173. SCHEFFE (1964) : *The Analysis of variance*, Wiley & Sons N. Y.
174. SHANNON C.E, WEAVER, W.E(1976) : La théorie mathématique de la communication, traduction française, J. Cosnier, Alii, Retz, CEPL.
175. SIMON GARY (1970) : Alternative analysis for the singly ordered contingency table, **JAS A**, Vol. 69, pp. 971-976.
176. STANKIEWICZ F.(1984) : Economie du chômage et de l'emploi, *Cujas*.
177. SUMMERS L. H(1986) : Why is the unemployment rate so very high near full employment ? **Brookings Papers on Economic Activity**, n° 2, pp. 339-396.
178. THELL H.(1967) : *Economics and information theory analysis*, North Holland Publishing Co, Amsterdam.
179. THELOT C.(1973) Mobilité professionnelle plus forte entre 1965 et 1970 qu'entre 1959 et 1964. **Eco & Stat.**, n° 51, pp. 3-33.
180. THELOT C.(1976) : Analyse statistique des migrations, *Annales de l'INSEE* n° 22/23 1976.
181. THIONET (1976) : Construction et reconstruction des tableaux statistiques, **Annales de l'INSEE**, n° 22/23.
182. TOMAS A. C.(1979) : Evolution des structures et disparités régionales de 1966 à 1975, **RERU** n° 2 pp. 237-258.

183. TOPEL R. H.(1986) : Local labor markets, **Jour. Polit. Econ.**, Vol 94, n° 3, pp. 111-143.
184. TROGNON P.(1979) : Les services marchands rendus principalement aux entreprises de 1960 à 1974 Collections de l'INSEE, E 59.
185. TURPIN E.(1981) : Disparités régionales, croissance et crise, **Eco & Stat**.
186. VALEYRE A.(1981) : Dynamique régionale de l'emploi et division spatiale du travail, CEE.
187. VELTZ P.(1983) : Les entreprises et la stabilisation de la main-d'œuvre, **RERU**, n° 1 pp. 27-41.
188. VERNIERES M.(1985) : L'emploi du tertiaire, *Economica*.
189. VIMONT C.(1981) : L'avenir de l'emploi, *Economica*.
190. VINCENS J.(1979) : Les nouveaux aspects du problème de l'emploi, **Revue d'Economie Politique**, n° 1.
191. VOLLE M. (1981) : Analyse des données , *Economica*.
192. WERMUTH N.(1976) : Analogies between multiplicative models in contingency tables and covariance selection, **Biometrics**, Vol. 32, pp 95-108
193. WILLIAMS D. A.(1976) : Improved likelihood ratio tests for complete contingency tables, **Biometrika**, Vol. 63 n° 1 pp. 33-37.
194. WILSON T.P.(1979) : On not interpreting coefficients.
195. ZAHN D. A., FEIN S. B.(1979) : Large contingency tables with large cells frequencies : a model search algorithm and alternative measures of fit, **Psychological Bulletin**, n°86, pp. 1189-1200.
196. ZIGHERA J. A. (1982) : Métiers et générations, **Eco & Stat**, n°145, pp. 19-27.
197. ZIGHERA J. A.(1985) : Partitioning information in multidimensional contingency tables and centering of log-linear parameters, **ASMAD**, Vol. 1, pp. 93-108.

TABLE DES MATIERES

| | |
|---|----|
| DEDICACES | |
| SOMMAIRE | |
| INTRODUCTION GENERALE..... | 1 |
|
 | |
| CHAPITRE I : RAPPELS THEORIQUES ET EMPIRIQUES DE L'EVOLUTION
GLOBALE DU MARCHE DU TRAVAIL EN FRANCE DE L'APRES GUERRE A LA
CRISE..... | 14 |
| 1.0- INTRODUCTION..... | 15 |
| 1.1- SCHEMAS D'ANALYSE THEORIQUE..... | 16 |
| 1.1.1- Cadre définitionnel du concept du marché..... | 16 |
| 1.1.1.1- Le marché, un lieu abstrait..... | 17 |
| 1.1.1.2- Le marché, un lieu géographique..... | 17 |
| 1.1.1.3- Les fonctions du marché..... | 18 |
| 1.1.2- Le concept de marché appliqué au travail..... | 19 |
| 1.1.2.1- Le travail : un objet d'échange ?..... | 19 |
| 1.1.2.1.1- L'approche néo-classique..... | 19 |
| 1.1.2.1.2- L'approche marxiste..... | 20 |
| 1.1.2.2- Récapitulatif des deux approches et situation particulière du marché du
travail..... | 21 |
| 1.1.3- Quelques modèles du marché du travail..... | 22 |
| 1.1.3.1- Modèles néo-classiques du marché du travail..... | 22 |
| 1.1.3.1.1- Hypothèses du modèle concurrentiel..... | 23 |
| 1.1.3.1.2- Limites des hypothèses du modèle..... | 23 |
| 1.1.3.1.3- Quelques apports et théories locales néo-classiques..... | 24 |
| 1.1.3.2- Modèle de segmentation du marché du travail..... | 27 |
| 1.1.3.2.1- Eléments généraux sur le processus de la segmentation..... | 27 |
| 1.1.3.2.2- Typologie du processus de la segmentation..... | 28 |
| 1.2 - EVOLUTION GLOBALE DU MARCHE DU TRAVAIL EN FRANCE
DE L'APRES-GUERRE A LA CRISE..... | 30 |
| 1.2.1- Approche globale du marché du travail..... | 31 |
| 1.2.1.1- Le marché du travail en France sur la période de forte croissance..... | 31 |
| 1.2.1.1.1- L'évolution quantitative du marché du travail..... | 31 |
| 1.2.1.1.2- Tendance d'homogénéisation du salariat..... | 32 |
| 1.2.1.1.3- Tendance d'hétérogénéisation du salariat..... | 32 |
| 1.2.1.2- Le marché du travail en France sur la période de crise..... | 34 |
| 1.2.1.2.1- Les traits marquants de la période de crise..... | 34 |
| 1.2.1.2.2- Eclatement du collectif du travail..... | 35 |
| 1.2.2- Dynamique régionale de l'emploi..... | 36 |
| 1.2.2.1- Redéploiement géographique des activités et structuration des régions
françaises..... | 36 |
| 1.2.2.2- Les grandes phases de la dynamique régionale de l'emploi..... | 38 |
| 1.2.2.2.1- La période 1954-1962..... | 39 |
| 1.2.2.2.2- La période 1962-1975..... | 39 |
| 1.2.2.2.3- La période de crise (à partir de 1975)..... | 40 |
| 1.2.2.2.4- Rapprochement des profils régionaux..... | 41 |

| | |
|---|----|
| CONCLUSION..... | 42 |
| CHAPITRE II : APPROCHE SPATIALISEE DU MARCHE DU TRAVAIL..... | 47 |
| 2.0- INTRODUCTION..... | 48 |
| 2.1- CONCEPT DE L'ESPACE ET COMPORTEMENTS DES AGENTS FACE A
L'ESPACE..... | 49 |
| 2.1.1- Notion de l'espace..... | 50 |
| 2.1.1.1- L'espace économique..... | 50 |
| 2.1.1.2- L'espace du marché du travail..... | 51 |
| 2.1.2- Rapport des agents économiques à l'espace..... | 53 |
| 2.1.2.1- Les rapports de l'individu à l'espace..... | 54 |
| 2.1.2.1.1- L'individu est sensible à son environnement..... | 54 |
| 2.1.2.1.2- L'interaction entre l'individu et l'espace..... | 55 |
| 2.1.2.2- Les rapports de l'entreprise à l'espace..... | 56 |
| 2.1.2.2.1- Eléments de la théorie traditionnelle de la localisation..... | 56 |
| 2.1.2.2.2- Remise en cause de la théorie traditionnelle de la localisation..... | 57 |
| 2.1.2.2.3- Différenciation spatiale et caractéristiques de la main-d'œuvre de
l'entreprise..... | 58 |
| 2.1.2.2.4- Différenciation spatiale et modalités d'aspiration de la main-d'œuvre..... | 60 |
| 2.1.2.3- Quelques éléments de synthèse..... | 61 |
| 2.2- INTEGRATION DE L'ESPACE DANS LE FONCTIONNEMENT DU
MARCHE DU TRAVAIL..... | 64 |
| 2.2.1- L'intérêt d'une localisation du marché du travail..... | 64 |
| 2.2.1.1- Diversité géographique des variables du marché du travail..... | 65 |
| 2.2.1.2- Renouveau du paradigme du développement régional..... | 66 |
| 2.2.1.3- Approche locale des politiques de l'emploi..... | 68 |
| 2.2.1.4- L'espace à travers les modèles macroéconomiques..... | 69 |
| 2.2.1.4.1- Modèles nationaux à répartition régionale..... | 70 |
| 2.2.1.4.2- Modèles régionaux à cohérence nationale..... | 71 |
| 2.2.2- L'espace dans le modèle traditionnel du marché du travail..... | 73 |
| 2.2.2.1- Spatialisation du marché concurrentiel sur la base des critères..... | 73 |
| 2.2.2.2- Spatialisation du marché concurrentiel sur la base d'une modification des
hypothèses du modèle..... | 75 |
| 2.2.2.2.1- Mise en évidence du facteur spatial dans l'hypothèse de la mobilité du
travail..... | 75 |
| 2.2.2.2.2- L'espace à travers les conditions de la diffusion de l'information..... | 77 |
| 2.2.2.2.3- Prise en compte de l'espace à travers l'hypothèse d'atomicité des agents..... | 78 |
| 2.3- ETUDE DE QUELQUES CAS DE MARCHE LOCALISE..... | 79 |
| 2.3.1- Exemples de marchés locaux..... | 79 |
| 2.3.1.1- Les bassins d'emploi..... | 80 |
| 2.3.1.2- Les zones d'emploi..... | 81 |
| 2.3.1.3- Marchés locaux concrets..... | 82 |
| 2.3.2- Détermination des limites géographiques du marché local..... | 83 |
| 2.3.2.1- Démarches empiriques..... | 83 |
| 2.3.2.2- Démarches statistiques..... | 85 |
| 2.3.2.3- Marché local dans la perspective de V. GIARD..... | 86 |

| | |
|---|-----|
| 2.3.2.3.1- Méthode fondée sur les migrations entre zones..... | 87 |
| 2.3.2.3.2- Méthode fondée sur les variables structurelles..... | 88 |
| 2.3.2.4- Eléments de synthèse et recherche d'un outil d'analyse approprié pour l'étude
des structures du marché..... | 89 |
| NOTES..... | 92 |
|
 | |
| CHAPITRE III : QUELQUES ELEMENTS DE BASE NECESSAIRES A L'ANALYSE DU
MODELE LOG-LINEAIRE..... | 95 |
|
 | |
| 3.0- INTRODUCTION..... | 96 |
| 3.1- RAPPEL SUR LES VARIABLES..... | 97 |
| 3.1.1- Eléments généraux..... | 98 |
| 3.1.2- Variables qualitatives..... | 98 |
| 3.1.2.1- Caractéristiques des variables qualitatives..... | 98 |
| 3.1.2.2- Recherche d'une méthodologie pour le traitement des variables qualitatives..... | 100 |
|
 | |
| 3.2- TABLEAU DE CONTINGENCE MULTIDIMENSIONNEL..... | 102 |
| 3.2.1- Définition d'un tableau de contingence..... | 102 |
| 3.2.2- Totaux marginaux et marges..... | 104 |
| 3.2.3- Marges impliquantes et impliquées..... | 105 |
|
 | |
| 3.3- QUELQUES PROCESSUS D'ECHANTILLONNAGE..... | 108 |
| 3.3.1- Processus d'échantillonnage Poissonnien..... | 108 |
| 3.3.2- Processus d'échantillonnage multinomial..... | 109 |
| 3.3.3- Processus d'échantillonnage produit-multinomial..... | 110 |
| 3.3.4- Famille exponentielle..... | 111 |
| 3.3.4.1- Définition d'une famille exponentielle..... | 112 |
| 3.3.4.2- Statistiques suffisantes..... | 113 |
|
 | |
| 3.4- ESTIMATION PAR LE PRINCIPE DU MAXIMUM DE VRAISEMBLANCE
EN GENERAL..... | 114 |
| 3.4.1- Problème d'estimation par le principe du maximum de vraisemblance..... | 115 |
| 3.4.2- Contraintes internes..... | 115 |
| 3.4.3- Contraintes externes..... | 117 |
| 3.4.4- Méthodes d'estimation en général..... | 118 |
| 3.4.4.1- Méthode d'estimation directe..... | 118 |
| 3.4.4.2- Approche par la procédure itérative..... | 119 |
|
 | |
| 3.5- PRESENTATION DE QUELQUES MODELES DE REGRESSION..... | 122 |
| 3.5.1- Modèle de régression linéaire..... | 123 |
| 3.5.1.1- Modèle de régression linéaire classique..... | 123 |
| 3.5.1.2- Modèle de régression polynomiale..... | 124 |
| 3.5.2- Modèle de régression multiplicative..... | 125 |
| 3.5.3- Modèle d'analyse de la variance..... | 127 |
| 3.5.4- Synthèse des modèles et situation du modèle log-linéaire..... | 128 |
|
 | |
| CONCLUSION..... | 129 |

CHAPITRE IV : MODELISATION LOG-LINEAIRE : FORMULATION DE BASE ET SON EXTENSION PAR LE BIAIS DU CENTRAGE DES PARAMETRES..... 130

| | |
|--|-----|
| 4.0- INTRODUCTION..... | 131 |
| 4.1- PRESENTATION DU MODELE LOG-LINEAIRE A TRAVERS LA
TREORIE DE L'INFORMATION..... | 132 |
| 4.1.1- Notion et signification des effets..... | 133 |
| 4.1.2- Notion d'information dans un tableau..... | 135 |
| 4.1.2.1- Quelques éléments de la théorie de l'information..... | 135 |
| 4.1.2.2- Quantité d'information dans un tableau quelconque..... | 136 |
| 4.1.2.3- L'information discriminante entre deux tableaux..... | 137 |
| 4.1.3- Approche introductive et forme structurelle du modèle Log-linéaire..... | 139 |
| 4.1.3.1- Exploration de la logique d'ensemble et introduction de la modélisation Log-
linéaire..... | 139 |
| 4.1.3.2- Forme structurelle du modèle Log-linéaire..... | 141 |
| 4.1.3.2.1- Modèles saturés..... | 142 |
| 4.1.3.2.2- Modèles non saturés..... | 144 |
| 4.1.4- Procédure traditionnelle de calcul des paramètres..... | 145 |
| 4.2- TECHNIQUES D'ESTIMATION POUR UN MODELE DONNE..... | 146 |
| 4.2.1- Recherche d'une configuration minimale suffisante..... | 147 |
| 4.2.2.1- Estimation fondée sur la méthode directe..... | 149 |
| 4.2.2.2- Estimation fondée sur la méthode indirecte..... | 150 |
| 4.2.2.2.1- L'algorithme d'ajustement proportionnel itératif..... | 151 |
| 4.2.2.2.2- L'algorithme d'ajustement de Newton-Raphson..... | 153 |
| 4.2.3- Autres méthodes d'estimation..... | 155 |
| 4.2.3.1 Méthode du score..... | 155 |
| 4.2.3.2 La méthode des moindres carrés pondérés..... | 156 |
| 4.2.3.3 La méthode du Khi-deux minimum..... | 156 |
| 4.3- TECHNIQUES DE CHOIX DES MODELES ET EXTENSION DU MODELE
LOG-LINEAIRE..... | 157 |
| 4.3.1- Techniques de choix entre des modèles Log-linéaires..... | 157 |
| 4.3.1.1- Statistiques de la qualité d'ajustement des modèles..... | 158 |
| 4.3.1.2- Choix de modèle fondé sur la statistique du G^2 | 159 |
| 4.3.1.2.1- Principes du choix entre deux modèles..... | 159 |
| 4.3.1.2.2- Limites des statistiques de X^2 et du G^2 | 160 |
| 4.3.1.2.3- Indices globaux d'ajustement des modèles..... | 161 |
| 4.3.1.3- Choix de modèle fondé sur la procédure Stepwise..... | 162 |
| 4.3.1.3.1- Approche ascendante..... | 163 |
| 4.3.1.3.1- Approche ascendante..... | 163 |
| 4.3.1.4- Choix de modèle fondé sur les termes standardisés des paramètres..... | 164 |
| 4.3.2- Extension du modèle Log-linéaire : son approche informationnelle..... | 165 |
| 4.3.2.1- Objet du centrage des paramètres..... | 165 |
| 4.3.2.2- Notion et propriétés des résidus..... | 167 |
| 4.3.2.3- Procédure de centrage des paramètres..... | 168 |
| 4.3.2.4- Décomposition de l'information entre les marges indépendantes..... | 170 |
| CONCLUSION..... | 171 |

| | |
|--|-----|
| CHAPITRE V : MODELISATION LOG-LINEAIRE DES TABLEAUX DE
CONTINGENCE AYANT DES MODALITES ORDINALES..... | 173 |
| 5.0- INTRODUCTION..... | 175 |
| 5.1- MESURE D'ASSOCIATION ET FORME STRUCTURELLE DU MODELE
LOG-LINEAIRE APPLIQUE A UN TABLEAU A VARIABLES ORDINALES..... | 175 |
| 5.1.1- Mesure générale de l'association..... | 176 |
| 5.1.2.- Modèles d'association fondés sur les ood-ratios..... | 177 |
| 5.1.2.1- Modèle d'association nulle..... | 178 |
| 5.1.2.2- Modèle d'association uniforme..... | 178 |
| 5.1.2.3.- Modèle d'association "ligne"..... | 179 |
| 5.1.2.4.- Modèle d'association "colonne"..... | 179 |
| 5.1.2.5- Modèle d'association "ligne-colonne"..... | 179 |
| 5.1.3.- Formulation Log-linéaire des modèles d'association..... | 180 |
| 5.1.3.1- Forme Log-linéaire des modèles d'association dans un tableau de dimension 2..... | 180 |
| 5.1.3.1.1- Modèle Log-linéaire appliqué à un tableau à variables nominale et ordinale..... | 180 |
| 5.1.3.1.2 Modèle Log-linéaire appliqué à un tableau où toutes les deux variables sont
ordinales..... | 184 |
| 5.1.3.2- Forme Log-linéaire des modèles d'association dans un tableau de dimension 3..... | 186 |
| 5.1.3.2.1- Modèle Log-linéaire appliqué à un tableau à une variable ordinale et deux
variables nominales..... | 186 |
| 5.1.3.2.2- Modèle Log-linéaire appliqué à un tableau à deux variables ordinales et une
variable nominale..... | 188 |
| 5.1.3.2.3- Modèle Log-linéaire appliqué à un tableau ayant trois variables ordinales..... | 189 |
| 5.2- AUTRES TYPES DE MODELES COMME OUTILS DE TRAITEMENT DES
TABLEAUX A VARIABLES ORDINALES..... | 191 |
| 5.2.1- Forme Logit des modèles d'association..... | 192 |
| 5.2.1.1- Formulation logit du modèle d'association dans un tableau de dimension..... | 193 |
| 5.2.1.1.1- Cas d'un tableau à une variable ordinale et une variable nominale..... | 193 |
| 5.2.1.1.2.- Cas d'un tableau à deux variables ordinales..... | 195 |
| 5.2.1.2.- Formulation logit du modèle d'association dans un tableau de dimension..... | 196 |
| 5.2.2- Une paramétrisation du modèle Log-linéaire fondée sur les polynôme
orthogonaux..... | 198 |
| 5.2.2.1- Forme générale du modèle..... | 199 |
| 5.2.2.2.- Quelques exemples de modèles illustratifs..... | 200 |
| 5.2.2.2.1- Cas d'un tableau 2x2..... | 200 |
| 5.2.2.2.2- Cas d'un tableau 3x3..... | 201 |
| 5.2.3- Forme Log-non linéaire des modèles d'association..... | 203 |
| 5.2.3.1- Caractéristiques du modèle Log-non linéaire..... | 204 |
| 5.2.3.1.1- Quelques généralités sur le modèle Log-non linéaire..... | 204 |
| 5.2.3.1.2- Propriétés du modèle Log-non linéaire..... | 206 |
| 5.2.3.2- Estimation des modèles d'association..... | 208 |
| 5.2.3.2.1- Estimation de la forme Log-linéaire des modèles d'association..... | 209 |
| 5.2.3.2.2- Estimation des modèles Log-non linéaires..... | 211 |
| 5.2.3.3- Ebauche d'une autre approche du modèle Log-non linéaire..... | 214 |
| 5.2.3.3.1- Généralités sur la méthode de l'analyse des correspondances..... | 214 |
| 5.2.3.3.2- Approche factorielle du modèle Log-non linéaire..... | 215 |
| CONCLUSION..... | 219 |

| | |
|---|-----|
| CHAPITRE VI : PRESENTATION DES DONNEES ET ANALYSE DES RESULTATS
STATISTIQUES..... | 221 |
| 6.0 INTRODUCTION | 222 |
| 6.1- PRESENTATION DES DONNEES D'ETUDE..... | 222 |
| 6.1.1- Réduction de la nomenclature de la variable profession | 224 |
| 6.1.2- Réduction de la nomenclature de la variable âge..... | 227 |
| 6.2.- EXPLORATION DE QUELQUES RESULTATS STATISTIQUES..... | 229 |
| 6.2.1- Sélection du meilleur modèle..... | 229 |
| 6.2.2- Effets de paramètres du modèle Log-linéaire..... | 235 |
| 6.2.2.1- Paramètres du modèle Log-linéaire de base..... | 235 |
| 6.2.2.2- Effets de paramètres centrés..... | 236 |
| 6.2.2.3- Effets de paramètres des modèles d'association..... | 239 |
| 6.3.- ANALYSE DES EFFETS SIMPLES..... | 244 |
| 6.3.- 1 Analyse de l'effet profession..... | 244 |
| 6.3.1.1- Profils différenciés des professions en niveau..... | 245 |
| 6.3.1.2- Profils différenciés des professions en évolution..... | 245 |
| 6.3.2.- Analyse de l'effet région..... | 248 |
| 6.3.2.1- Profils différenciés des régions..... | 248 |
| 6.3.2.2.- Profils différenciés des régions un évolution..... | 248 |
| 6.3-3 -Analyse des effets âge et sexe..... | 250 |
| 6.3.3.1- Analyse de l'effet âge..... | 250 |
| 6.3-3.2- Analyse de l'effet sexe..... | 251 |
| CONCLUSION..... | 252 |
| ANNEXES..... | 253 |
| CHAPITRE VII : ANALYSE STRUCTURELLE ET DYNAMIQUE DE L'EMPLOI
SALARIE ENTRE 1976 ET 1981..... | 261 |
| 7.0 INTRODUCTION | 262 |
| 7.1 ANALYSE REGIONALE DE LA STRUCTURE DES PROFESSIONS..... | 264 |
| 7.1.1- Profils différenciés des professions et des régions..... | 265 |
| 7.1.2- Analyse détaillée des professions dans un groupe de régions typées..... | 268 |
| 7.1.2.1- Ile-de-France..... | 269 |
| 7.1.2.2- Rhône – Alpes..... | 270 |
| 7.1.2.3- Provence Alpes - Côte d' Azur – Corse..... | 272 |
| 7.1.2.4- Franche Comté..... | 274 |
| 7.1.2.5- Bretagne..... | 275 |
| 7.1.3- Déterminants de la localisation du tertiaire..... | 277 |
| 7.1.3.1- Facteurs extra-économiques..... | 278 |
| 7.1.3.2- Facteurs conjoncturels..... | 279 |
| 7.2. PARTICIPATION FEMININE PROFESSIONNELLE ET REGIONALE..... | 280 |

| | |
|---|-----|
| 7.2.1- Participation féminine professionnelle..... | 281 |
| 7.2.1.1- Typologie d'ensemble : une polarisation de la participation féminine..... | 283 |
| 7.2.1.2- Evolution de la participation féminine professionnelle sur la période 1976 –
1981..... | 283 |
| 7.2.2- Participation féminine régionale..... | 286 |
| 7.2.2.1- Participation féminine : des comportements régionaux relativement diversifiés..... | 287 |
| 7.2.2.2- Evolution régionale de la participation féminine entre 1976 et 1981..... | 289 |
| 7.2.2.3- Rôle déterminant des métiers dans la féminisation régionale..... | 291 |
|
 | |
| 7.3- ANALYSE DE LA STRUCTURATION DES METIERS ET REGIONS PAR
LE FACTEUR AGE..... | 294 |
| 7.3.1- L'âge comme facteur de structuration des salariés suivant les métiers..... | 294 |
| 7.3.2- Main-d'oeuvre juvénile et son rôle spécifique : sur le marché de travail..... | 298 |
| 7.3.3- L'âge comme facteur de structuration des salariés suivant les régions..... | 303 |
| 7.3.3.1- Une opposition Nord-Sud : le Nord "jeune" et le Sud "vieux"..... | 303 |
| 7.3.3.2- Autres déterminants de la fracturation spatiale des salariés suivant les classes
d'âge..... | 307 |
| 7.3.3.3- Typologie régionale des salariés selon les tranches d'âge..... | 310 |
|
 | |
| CONCLUSION..... | 314 |
| ANNEXES..... | 320 |
| CONCLUSION GENERALE..... | 335 |
| ANNEXES..... | 342 |
| BIBLIOGRAPHIE..... | 368 |
| TABLE DE MATIERES..... | 380 |