

DU GROUPE À L'INDIVIDU : L'EXEMPLE DES COMPORTEMENTS MIGRATOIRES*

En se donnant pour objet l'étude du renouvellement des populations, la démographie s'est délibérément placée dans un cadre qui privilégie l'analyse de grandeurs agrégées : le « stock » d'individus est modifié, en volume et en structure, par le « flux » des naissances, décès et tous autres événements intervenant dans le mouvement de population. Mais la compréhension de ces flux renvoie à des comportements qui font apparaître sous la statistique, l'individu ; le recours croissant aux données d'enquêtes fondées sur des reconstitutions biographiques met en lumière, plus clairement que jamais, l'existence d'une micro-démographie à côté ou en complément de la macro-démographie.

*Ce dualisme n'est pas propre à la discipline ; il affecte toutes les sciences sociales. Par exemple, évoquant la coexistence d'une micro-économie et d'une macro-économie, Marc Blaug** souligne qu'elle crée « une sorte de schizophrénie intellectuelle dans laquelle les techniques de l'une ne recouvrent pas le domaine de l'autre. Cette situation étant insatisfaisante, des économistes ont essayé de jeter un pont entre le consommateur et la fonction de consommation globale ou entre l'investisseur et les facteurs de l'agrégat-investissement. Le pont reste cependant inachevé et l'étudiant en économie doit se préparer à utiliser deux boîtes à outils ».*

*Daniel COURGEAU*** apporte ici sa contribution à la construction du pont en mettant en parallèle l'analyse des données individuelles et des données agrégées sur les migrations.*

Selon que les sciences humaines abordent les comportements des groupes ou ceux des individus, les hypothèses, les objectifs, les formulations et les caractéristiques considérés vont être très différents. Nous envisageons ici la modélisation de ces comportements, qui permet de mettre au clair les problèmes rencontrés lorsque l'on passe d'un niveau à l'autre.

* L'auteur tient à remercier ici Nico Keilman (Central Bureau of Statistics, Norvège) et Nicole Tabard (CNRS) pour leurs commentaires sur une première version de cet article.

** Mark BLAUG, « Economics », in *Encyclopedia Universalis*, Macropedia, vol. 6, p. 270, éd. 1973-1974.

*** INED.

Nous présentons d'abord ces problèmes sous forme générale, en les éclairant ensuite par l'exemple des comportements migratoires.

La première approche se situe au *niveau agrégé* et cherche à définir des caractéristiques globales du groupe en vue de comprendre son comportement. Elle va donc essayer de représenter une réalité complexe par un schéma simplifié, qui articulera les grandes caractéristiques du groupe entre elles. Bien plus, cette approche sera souvent basée sur des comportements individuels hypothétiques, que l'on ne pourra vérifier du fait que les données qu'elle utilise sont agrégées. Ainsi, par exemple, les modèles de migration vont expliquer les flux de migrants sous l'hypothèse que leur comportement est influencé par diverses caractéristiques des zones de départ et d'accueil et par la distance physique ou sociale qui sépare ces zones. Cette approche est suivie depuis fort longtemps [Young, 1924] et se situe d'emblée à un niveau macro-géographique : seules les caractéristiques des zones influenceront sur les déplacements des individus. De la même façon les chercheurs qui travaillent dans le domaine de l'analyse structurelle, contextuelle, écologique, etc., reconnaîtront là leur problématique.

La seconde approche se situe au *niveau individuel* et va travailler, non plus sur des caractéristiques globales, mais sur les divers éléments de la biographie de chaque personne pour expliquer son comportement. Les modèles correspondants se sont développés récemment en sociologie [Tuma et Hannan, 1984], en économie [Lancaster, 1990] et en démographie [Courgeau et Lelièvre, 1989]. Ainsi, par exemple, ces modèles vont expliquer, non plus les flux, mais les comportements des migrants [Courgeau, 1976 ; Sandefur et Scott, 1981] : ils se situent alors à un niveau *micro-géographique*. Ils utilisent pour ce faire des données d'enquêtes biographiques qui recueillent les divers événements de leur vie professionnelle, migratoire, etc. Ils vont alors relier la probabilité qu'un individu connaisse un événement donné, une migration par exemple, à son histoire passée, à ses caractéristiques et à l'information dont il dispose sur les diverses zones d'accueil.

L'intégration de ces deux approches est rendue difficile du fait que la première prédit un comportement collectif à l'aide de *caractéristiques du groupe*, tandis que la seconde prédit un comportement individuel à l'aide de *caractéristiques biographiques*. Une difficulté supplémentaire apparaît du fait que la première approche travaille en *transversal* tandis que la seconde travaille en *longitudinal* : dans le cas des migrations on prévoit en transversal un flux de migrants, tandis qu'en longitudinal il s'agit d'une probabilité de migrer tout au long de l'existence. Pour ces raisons, les tentatives d'intégration des deux approches sont encore très restreintes [Sanders, 1992, pp. 51-56].

Nous allons essayer ici de relier le groupe à l'individu dans le cadre des modèles de migration. Il nous est en effet apparu nécessaire de préciser au maximum les conditions sous lesquelles la comparaison s'opère, pour entamer ce travail délicat. Nous pensons cependant que les résultats obtenus

pourront être généralisés au problème plus global que nous avons posé dans cette introduction.

Pour voir plus en détail comment les deux niveaux sont imbriqués, nous préciserons d'abord les modèles les plus souvent utilisés pour chacun d'eux. Nous les simplifierons ensuite, pour ne travailler dans un premier temps que sur des flux d'émigrants. Nous reformulerons ensuite ces modèles en utilisant des variables identiques pour les deux niveaux et comparerons théoriquement les résultats auxquels ils conduisent. Nous appliquerons ensuite ces formulations aux données de l'enquête sur la « biographie familiale, professionnelle et migratoire », dite « 3B ». Cela nous permettra de préciser plus avant les relations entre les deux niveaux d'agrégation, avant de conclure.

I. – Formulation classique des deux approches

Nous présentons d'abord les deux types de modèles en définissant avec précision leurs objectifs et leur formulation. Nous n'essayons pas encore de relier entre elles les caractéristiques utilisées.

Analyse au niveau agrégé Cette analyse va expliquer les flux de migrants entre zones à l'aide de diverses caractéristiques, par des modèles de régression multiple. Ces modèles peuvent être formalisés de diverses façons ; ils sont le plus souvent multiplicatifs, mais peuvent encore être additifs ou plus complexes [Stillwell, 1975].

Bien que l'approche soit de nature macro-géographique, les caractéristiques sélectionnées dans les modèles le sont en fonction de l'idée que l'on se fait d'un comportement migratoire individuel [Greenwood, 1975 ; Puig, 1981]. On les distingue souvent en trois grands types.

Le premier type mesure le pouvoir de répulsion de la zone de départ ou, au contraire, des raisons d'y rester : taux de chômage, niveau des salaires, part d'individus dans un niveau d'éducation donné, etc. Ainsi on peut supposer qu'un taux de chômage important et un bas niveau de salaire poussent à l'émigration. Le second type va représenter le pouvoir d'attraction de la zone de destination ou, au contraire, les raisons de ne pas y aller. Les caractéristiques peuvent être les mêmes que les précédentes, mesurées maintenant sur la zone de destination avec un effet inverse du précédent. Ainsi un taux de chômage faible et un haut niveau de salaire peuvent rendre une région attractive. Enfin un troisième type de variables va mesurer l'interaction qui existe entre zone de départ et d'arrivée. Divers types de distances ont été proposés pour ce faire : distance physique, distance mesurée par les postes offerts entre les deux zones [Stouffer, 1940], etc.

Le choix de ces caractéristiques et le contenu explicatif qui leur est donné montrent clairement qu'il y a un modèle de comportement individuel

sous-jacent. D'ailleurs l'interprétation des effets de ces caractéristiques est souvent donné en termes personnels : on dira que c'est l'attrait d'un salaire plus élevé qui pousse un individu à migrer, alors que le modèle montre seulement une relation entre effectif de migrants et salaire moyen de la zone de destination.

La formulation mathématique générale de ces modèles peut s'écrire :

$$M_{ij} = f(X_i, Y_j, I_{ij}, \varepsilon_{ij}) \quad [1]$$

où M_{ij} est le flux observé entre i et j , f une fonction dont la forme peut varier, X_i et Y_j des vecteurs de caractéristiques de la zone de départ et d'arrivée, I_{ij} des mesures de l'interaction entre ces zones et ε_{ij} le résidu non expliqué dépendant de toutes les autres caractéristiques non prises en compte.

Ces modèles sont souvent estimés sur des flux de migrants, mesurés par une question censitaire sur le lieu de résidence à une date antérieure. Les diverses caractéristiques peuvent être fournies par un recensement antérieur ou être issues de différentes sources (déclaration annuelle de salaire, etc.). Ces caractéristiques devraient normalement être mesurées en début de période, mais l'absence fréquente de données à cette date conduit de nombreux auteurs à lever cette hypothèse [Puig, 1981].

Analyse au niveau individuel

Cette analyse suppose que l'individu, α , effectue des migrations successives tout au long de sa vie entre r régions du territoire. Notons T_k^α la durée séparant la $(k-1)^{\text{ième}}$ migration de la $k^{\text{ième}}$, I_k^α la région de résidence pendant cette période. Il s'agit de variables aléatoires, auxquelles la théorie des probabilités s'applique. On peut ainsi modéliser le quotient instantané de rang k , défini comme la probabilité d'effectuer une migration de i vers j entre t et $t+dt$, conditionnée par le séjour en i jusqu'à la date t , divisée par dt :

$$m_{ijk}^\alpha(t) = \lim_{dt \rightarrow 0} \frac{P(T_k^\alpha < t + dt, I_{k+1}^\alpha = j \mid T_k^\alpha \geq t, I_k^\alpha = i)}{dt} \quad [2]$$

Il est ensuite possible de faire intervenir l'effet sur ce quotient de nombreuses caractéristiques individuelles, mesurées en t ou auparavant. Ces caractéristiques sont souvent obtenues par une enquête rétrospective retraçant les diverses histoires de vie : activité professionnelle, niveau d'éducation, statut d'occupation du logement, etc. [Courgeau et Lelièvre, 1989]. Ces caractéristiques sont souvent mesurées par des variables binaires, valant 1 si les individus les possèdent et 0 sinon. On peut également faire intervenir des variables discrètes (salaire, nombre d'enfants, etc.) ou même plus complexes. Il est enfin possible, à l'aide de ces enquêtes, de disposer de certaines informations de l'individu sur les zones de migration possibles : lieux où il a antérieurement résidé, localisation de parents ou d'amis,

etc. Il est dès lors possible d'écrire le quotient instantané de migration de l'individu α , sous la forme générale suivante :

$$m_{ijk}^{\alpha}(t; x_i^{\alpha}(t), y_j^{\alpha}(t)) \quad [3]$$

où $x_i^{\alpha}(t)$ correspond à ses caractéristiques dans la zone i et $y_j^{\alpha}(t)$ à l'information qu'il possède sur j . La formulation plus précise d'un tel quotient peut se faire sous forme entièrement paramétrique [Courgeau, 1985a], soit sous forme semi-paramétrique [Courgeau et Lelièvre, 1986]. Dans le premier cas, les paramètres sont estimés par la méthode du maximum de vraisemblance, tandis que dans le second on doit utiliser une vraisemblance partielle [Cox, 1972]. On dispose simultanément d'estimations des variances et des covariances entre les divers paramètres.

II. – Rapprochement des deux formulations

Pour relier ces deux modèles, des hypothèses supplémentaires sont nécessaires. Il nous faut d'abord introduire un temps commun, puis simplifier les formules précédentes et enfin expliciter les modèles utilisés. Ce n'est qu'une fois ce travail terminé que nous pourrons voir les relations qui existent entre eux.

Un temps commun pour des modèles simplifiés

Les modèles présentés plus haut utilisent une temporalité différente. Le modèle au niveau agrégé s'applique en transversal sur une période de temps; le modèle au niveau individuel s'applique en longitudinal sur des intervalles de temps entre migrations successives.

Nous allons envisager ici une courte durée de temps, un an par exemple ($t_0, t_0 + 1$), au cours de laquelle nous mesurons le flux de migration M_{ij} et le quotient instantané m_{ij} , supposé constant au cours de cette durée. Nous n'avons donc plus à faire intervenir le rang de la migration, k , mais en revanche la date de début, t_{α} , de la période considérée doit être connue. Sa durée de séjour à l'instant t_0 sera $(t_0 - t_{\alpha})$.

Nous simplifions ensuite les modèles en travaillant sur les flux d'émigration des zones, laissant de côté ici les flux détaillés qui existent entre elles. Dans ce cas il n'y a plus lieu de faire intervenir les caractéristiques Y et I de la formule [1] ni les caractéristiques $y(t)$ de la formule [3], dont l'effet sur les flux d'émigration est supposé négligeable devant celui des caractéristiques de la zone de départ.

Précisons maintenant le modèle que nous allons utiliser pour les données individuelles. Nous avons choisi un modèle semi-paramétrique, qui évite de spécifier le quotient sous-jacent. Si nous travaillons en temps discret ce modèle s'explique sous la forme suivante [Kalbfleisch et Prentice, 1980, p. 36], en fonction de la durée de séjour $(t_0 - t_{\alpha})$:

$$m_{i.}(t_0 - t_\alpha; x_i^\alpha) = 1 - [1 - m_{i.}^0(t_0 - t_\alpha)]^{\exp x_i^\alpha \beta} = 1 - [1 - m_{i.}^0(t_0 - t_\alpha)]^{\exp \sum_{k=1}^n \beta_k x_{ik}^\alpha} \quad [4]$$

où $m_{i.}^0(\cdot)$ est un quotient sous-jacent identique pour tous les individus, mais dépendant de la durée de séjour de chacun ($t_0 - t_\alpha$), x_i^α est un vecteur ligne de caractéristiques individuelles (x_{ik}^α) et β un vecteur colonne de paramètres à estimer (β_k). Ces paramètres sont censés ici être indépendants des zones dans lesquelles les individus vivent. Pour simplifier cette relation nous allons supposer que les quotients sous-jacents sont de faible valeur, ce qui permet d'écrire plus simplement :

$$m_{i.}(t_0 - t_\alpha; x_i^\alpha) \approx m_{i.}^0(t_0 - t_\alpha) \exp \sum_{k=1}^n \beta_k x_{ik}^\alpha \quad [5]$$

Lorsque les variables sont binaires, on peut interpréter $\exp \beta_k$ comme le risque relatif d'un individu ayant la caractéristique k par rapport à ceux qui ne l'ont pas. Nous verrons plus loin comment estimer ces paramètres β .

Les modèles agrégés usuellement utilisés sont de type multiplicatif [Stillwell, 1975]. Pour être cohérents avec le modèle choisi au niveau individuel, nous allons utiliser un modèle multiplicatif de type exponentiel :

$$\frac{M_i}{N_i} = \exp(X_i \gamma) = \exp\left(\gamma_0 + \sum_{k=1}^n \gamma_k X_{ik}\right) \quad [6]$$

où M_i/N_i est le taux d'émigration de la zone i , X_i le vecteur ligne des diverses caractéristiques « macro » de cette zone (X_{ik}) et γ le vecteur colonne correspondant à leur effet (γ_k). Si l'on travaille sur des variables non centrées, il est nécessaire de faire intervenir un terme constant, γ_0 , dans cette régression. On peut estimer ces divers paramètres en passant aux logarithmes, ce qui rend le modèle linéaire.

Relations théoriques entre les deux modèles

zone i :

A partir des probabilités individuelles d'émigrer on estime l'espérance mathématique du nombre annuel d'émigrants de la

$$E(M_{i.}) \approx \sum_{\alpha=1}^{N_i} m_{i.}^0(t_0 - t_\alpha) \exp x_i^\alpha \beta = \sum_{\alpha=1}^{N_i} m_{i.}^0(t_0 - t_\alpha) \exp \sum_{k=1}^n \beta_k x_{ik}^\alpha \quad [7]$$

Rappelons que les caractéristiques individuelles, x_{ik}^α sont mesurées ici par des variables binaires ou des variables discrètes. On peut donc calculer les valeurs moyennes de ces caractéristiques individuelles observées dans la zone i :

$$X_{ik} = E(x_{ik}^\alpha) = \frac{\sum_{\alpha=1}^{N_i} x_{ik}^\alpha}{N_i} \quad [8]$$

On relie ainsi les caractéristiques individuelles aux caractéristiques des zones dans lesquelles ces individus vivent. On peut donc réécrire le modèle au niveau agrégé sous la forme suivante :

$$\frac{M_i}{N_i} = \exp\left(\gamma_0 + \sum_{k=1}^n \left[\gamma_k \frac{\sum_{\alpha=1}^{N_i} x_{ik}^\alpha}{N_i}\right]\right) = \exp\left(\gamma_0 + \frac{1}{N_i} \sum_{\alpha=1}^{N_i} \sum_{k=1}^n \gamma_k x_{ik}^\alpha\right) \quad [9]$$

La dernière relation a été obtenue en intervertissant l'ordre des sommations sur les caractéristiques (k) et les individus (α).

La comparaison des deux effectifs de migrants estimés par la formule [7], au niveau individuel, et [9], au niveau agrégé, devrait nous permettre de voir s'il existe des relations simples entre les paramètres. Or on voit, que si l'on suppose que $\beta_k = \gamma_k$, et que l'on note plus simplement :

$$e_i^\alpha = \exp \sum_{k=1}^n \beta_k x_{ik}^\alpha = \exp \sum_{k=1}^n \gamma_k x_{ik}^\alpha \quad [10]$$

on arrive alors à la relation plus simple :

$$\sum_{\alpha=1}^{N_i} m_i^0 (t_0 - t_\alpha) e_i^\alpha = N_i (\exp \gamma_0) \left(\prod_{\alpha=1}^{N_i} e_i^\alpha \right)^{\frac{1}{N_i}} \quad [11]$$

En notant que $m_i^0 (t_0 - t_\alpha)$ ne contient pratiquement aucune information sur les e_i^α [Cox, 1972], on peut les considérer comme deux variables aléatoires indépendantes. On peut dans ce cas réécrire cette relation :

$$\bar{m}_i^0 \sum_{\alpha=1}^{N_i} e_i^\alpha = (\exp \gamma_0) \left(\prod_{\alpha=1}^{N_i} e_i^\alpha \right)^{\frac{1}{N_i}} \quad [12]$$

où \bar{m}_i^0 est la somme des valeurs $m_i^0 (t_0 - t_\alpha)$ prise sur tous les individus de la zone i . On voit alors que dès que la population N_i comporte plus d'un individu, l'hypothèse $\beta_k = \gamma_k$ ne tient pas. Il n'est donc pas possible de relier simplement les paramètres estimés au niveau « micro » sur les caractéristiques individuelles, aux paramètres estimés au niveau « macro » sur les caractéristiques des zones.

En revanche, si l'on fait intervenir au niveau individuel les caractéristiques des zones sur la probabilité de migrer, la situation va changer. On a déjà indiqué qu'au niveau agrégé, on supposait un modèle de comportement qui rendait l'individu sensible aux caractéristiques des zones :

un chômage important dans une région est supposé pousser un individu actif ou chômeur vers une région où ce chômage est moindre. Dans ces conditions on peut réécrire la relation [7] en faisant intervenir les caractéristiques des zones plutôt que les caractéristiques individuelles :

$$\sum_{\alpha=1}^{N_i} m_i^0 (t_0 - t_\alpha) \exp \left(\sum_{k=1}^n \beta_k \frac{\sum_{i=1}^{N_i} x_{ik}^\alpha}{N_i} \right) =$$

[13]

$$= \sum_{\alpha=1}^{N_i} m_i^0 (t_0 - t_\alpha) \exp \left(\frac{1}{N_i} \sum_{\alpha=1}^{N_i} \sum_{k=1}^n \beta_k x_{ik}^\alpha \right)$$

On voit facilement apparaître dans ce cas les conditions d'identité des formules [9] et [13], quelles que soient les caractéristiques des individus :

$$\begin{cases} \beta_k = \gamma_k \\ E(m_i^0 (t_0 - t_\alpha)) = \exp \gamma_0 \end{cases} \quad [14]$$

Cette dernière relation suppose à nouveau que $m_i^0 (t_0 - t_\alpha)$ ne contient aucune information sur les paramètres β .

Il peut enfin être intéressant de faire intervenir simultanément les caractéristiques des individus et des zones où ils vivent, dans un même modèle au niveau individuel. La comparaison des paramètres obtenus pour les caractéristiques « macro » et « micro » devrait permettre d'éclairer les liens entre leurs effets sur la migration.

III. – Test des divers modèles sur les données de l'enquête «3B»

Il nous faut maintenant utiliser un fichier issu d'une même enquête pour comparer empiriquement les résultats auxquels les divers modèles conduisent. Définissons d'abord plus précisément le fichier sur lequel nous allons travailler en indiquant ses faiblesses pour réaliser un tel test.

Échantillon et migrations considérées

L'enquête «3B» a recueilli les biographies familiale, professionnelle et migratoire d'un échantillon de 4 602 individus, âgés de 45 à 69 ans en 1981. Pour faire une analyse statistique précise, il aurait été nécessaire de décomposer cette population en générations d'un sexe donné, dont on observe les départs des diverses régions françaises au cours d'une même année. La taille de l'échantillon nous empêche d'entrer dans un tel détail.

Nous nous contenterons ici de travailler sur toutes les générations observées en distinguant cependant les comportements selon le sexe. En effet, des différences de comportement entre hommes et femmes, vis-à-vis de la migration, sont très importantes [Courgeau, 1985b]. Nous travaillerons pour réaliser ce test sur la population féminine, la plus nombreuse (2 552 individus). Nous considérons ici l'émigration des régions françaises. Cependant, étant donné les faibles effectifs observés nous avons dû regrouper certaines Régions de Programme : le tableau 1 donne la composition et la liste des régions retenues.

TABLEAU 1. – LISTE DES RÉGIONS RETENUES POUR L'ÉTUDE AVEC LES RÉGIONS DE PROGRAMME CORRESPONDANTES

Régions	Code
Ile-de-France	1
Champagne-Ardennes et Picardie	2
Haute-Normandie et Basse-Normandie	3
Centre	4
Bourgogne	5
Nord-Pas-de-Calais	6
Lorraine	7
Alsace et Franche-Comté	8
Pays de la Loire	9
Bretagne	10
Poitou-Charentes	11
Aquitaine	12
Midi-Pyrénées	13
Rhône-Alpes	14
Limousin et Auvergne	15
Languedoc-Roussillon	16
Provence-Alpes-Côte d'Azur et Corse	17

Toujours à cause de la faiblesse des effectifs nous n'avons pas pu travailler sur une courte période d'un an. Nous avons dû regrouper les migrations effectuées de 1950 à 1960, en considérant les premiers changements de région réalisés au cours de cette période, à partir du lieu de résidence en 1950. Nous avons, bien entendu, éliminé les femmes résidant à l'étranger en 1950 (228), les conditions d'une migration internationale étant très différentes de celles d'une migration interne.

A ce manque de précision sur la cohorte observée et sur la période de migration, va s'ajouter le problème du choix des caractéristiques considérées. A partir d'analyses antérieures des changements de départements [Courgeau, 1985] nous avons considéré cinq caractéristiques des femmes en 1950. Toujours à cause des faibles effectifs observés nous ne pouvons pas tester l'effet d'un plus grand nombre de variables. La première de ces caractéristiques est le statut matrimonial : nous faisons l'hypothèse d'une réduction de la mobilité des femmes mariées. La seconde et la troisième sont liées au statut d'occupation du logement : logées chez les parents ou propriétaires. Nous supposons ici que les femmes logées chez leurs parents

ont une mobilité plus faible que les locataires ou les femmes logées gratuitement, prises comme groupe témoin, mais cependant plus forte que celle des propriétaires. La quatrième caractéristique est liée à l'activité : nous supposons que les inactives auront une mobilité facilitée par rapport à celle des actives. Enfin la cinquième caractéristique est professionnelle : une agricultrice ou une salariée agricole devraient avoir une mobilité bien moindre de celle des autres actives. On va simultanément chercher à voir si les caractéristiques « macro » correspondantes ont un effet similaire sur les flux de migrants, en calculant les pourcentages de femmes ayant ces diverses caractéristiques dans chacune des zones.

Il faut bien voir ici que les résultats régionaux auxquels nous arriverons, ne sont pas à considérer en tant que tels, mais en tant qu'éléments parfaitement comparables entre eux, puisqu'élaborés à partir de la même source. Notre souci est ici de comparer les différents niveaux d'agrégation et non pas de donner une vue précise de l'émigration régionale : la taille de l'échantillon ne le permet pas.

Estimation des paramètres et reformulation des modèles

Nous travaillons maintenant sur la période s'étendant de 1950 à 1960 et observons les femmes n'ayant pas connu de migration entre la date de leur dernière installation, qui peut être leur date de naissance, et l'année 1950. L'échantillon est dit « tronqué à gauche », car échappent à notre observation les migrations faites avant 1950. L'utilisation de telles données pose de nombreux problèmes et nécessite des hypothèses [Courgeau et Lelièvre, 1989, pp. 52 à 56]. Cependant, si l'on suppose que la probabilité de migrer ne dépend que de la durée écoulée depuis la dernière installation et des caractéristiques de chaque individu en 1950, nous avons une possibilité d'estimer les β et les $m_i^0(t_0 - t_\alpha)$.

Supposons, en effet, que l'individu α fasse une migration après une durée $(t - t_\alpha)$ passée dans la région d'origine. La population soumise au risque d'effectuer cette migration sera composée de tous les individus pour lesquels la durée de séjour est supérieure ou égale à $(t - t_\alpha)$ et qui ont été observés en 1950 avec une durée de séjour inférieure ou égale à $(t - t_\alpha)$. On voit dès lors en quoi ce modèle diffère des modèles classiques d'analyse des biographies : la population soumise au risque ne va pas décroître avec la durée de séjour, comme dans le modèle classique, mais au contraire va commencer par croître avec cette durée de séjour, au moins tant que celle-ci n'est pas trop élevée.

En dépit de cela, on peut toujours calculer la probabilité de migrer de cet individu, conditionnellement à la population soumise au risque :

$$\frac{m_i^0(t_0 - t_\alpha) \exp x_i^\alpha \beta}{\sum_{l \supset R_i} m_i^0(t_0 - t_\alpha) \exp x_i^l \beta} \quad [15]$$

où R_i est l'ensemble des individus soumis au risque après 1950 et ayant passé $(t - t_\alpha)$ années dans la région d'origine. On voit que cela conduit à une vraisemblance partielle, du type de celle proposée par Cox [1972] puisqu'on peut simplifier le numérateur et le dénominateur par $m_i^0(t_0 - t_\alpha)$. Cependant du fait que la population soumise au risque n'est pas uniformément décroissante, les logiciels classiques ne permettent pas d'estimer les paramètres β .

Pour ce faire nous avons utilisé un modèle à épisodes multiples faisant intervenir la durée écoulée depuis un instant initial unique pour tous les épisodes [Hamerle, 1989]. Le premier épisode, sans intérêt pour nous, se termine en 1950. Le second épisode que nous utilisons ici pour notre estimation débute en 1950 et dure jusqu'à la migration suivante. Bien entendu, si elle ne se produit pas avant 1961 on observe un intervalle ouvert. Cette méthode permet d'estimer correctement les paramètres β et la fonction $m_i^0(t_0 - t_\alpha)^{(1)}$.

Les paramètres du modèle au niveau agrégé s'estiment, comme précédemment, par une régression linéaire multiple, en passant aux logarithmes.

Du fait que l'on travaille sur une durée de 10 ans la formule [7] doit être modifiée. La probabilité que l'individu migre entre 1950 et 1960 doit s'estimer en fonction des durées de séjour. On peut dans ce cas écrire le nombre de migrants attendus :

$$E(M_i) = \sum_{\alpha=1}^{N_i} [S^0(1950 - t_\alpha)^{\exp x_i^\alpha \beta} - S^0(1960 - t_\alpha)^{\exp x_i^\alpha \beta}] \quad [16]$$

où $S^0(t)$ correspond à la fonction de séjour dans une région.

La formule [9], en revanche, reste inchangée. Il en résulte que les liens entre β et γ seront plus compliqués que ceux donnés par la formule [14]. L'application pratique va cependant montrer qu'ils restent toujours à peu près linéairement liés.

Résultats obtenus Nous allons d'abord tester la proximité des résultats obtenus au niveau individuel et agrégé à l'aide des caractéristiques « macro » des diverses zones. Le tableau 2 porte ces résultats. On voit que quatre caractéristiques ont un effet significatif sur les probabilités d'émigration des zones au niveau individuel, contre trois seulement au niveau agrégé. Les proportions de mariées et de femmes logées chez les parents réduisent cette probabilité : cela correspond aux hypothèses faites précédemment sur les caractéristiques individuelles. Également la proportion d'inactives augmente cette probabilité. En revanche la proportion d'agricultrices joue en sens inverse de ce que l'on attendait : nous reviendrons plus loin sur ce résultat.

(1) Nous avons utilisé pour ce faire le programme TDA, rédigé par Gotz Rohwer.

TABLEAU 2. — ESTIMATION DES PARAMÈTRES DES MODÈLES « MICRO », « MACRO » ET DE LEUR ÉCART-TYPE, EN FAISANT INTERVENIR LES DIVERSES CARACTÉRISTIQUES MESURÉES AU NIVEAU « MACRO »

Caractéristique	Modèle « micro »		Modèle « macro »	
	Paramètre estimé	Écart-type	Paramètre estimé	Écart-type
Proportion de mariées	- 6,987***	1,910	- 5,879***	1,920
Proportion de logées chez les parents	- 3,107**	1,434	- 2,404	1,452
Proportion d'inactives	5,472***	1,830	4,706**	1,934
Proportion d'agricultrices	3,468***	0,788	2,897**	0,980
Constante	-		0,936	1,365
Maximum du log. de vraisemblance	- 17 886,12			
R ²	-		0,514	

*** Résultat significatif au seuil de 1 %
** Résultat significatif au seuil de 5 %.

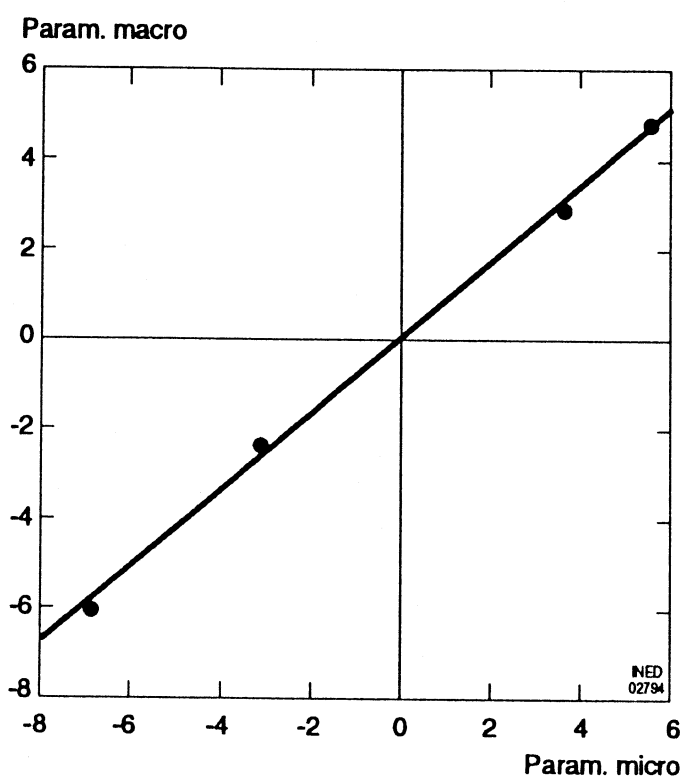


Figure 1 — Relations entre paramètres correspondant aux caractéristiques des diverses zones, estimés au niveau individuel (micro) et au niveau agrégé (macro)

La figure 1 montre que ces divers paramètres sont linéairement reliés entre eux, mais que les estimations au niveau individuel sont toujours plus fortes, en valeur absolue, que celles faites au niveau agrégé. On peut, dès lors, confirmer le résultat théorique avancé précédemment : si l'égalité des paramètres n'est plus vérifiée lorsque l'on travaille sur une période de longue durée, la proportionnalité entre eux reste correcte. On constate également que les écarts types de ces estimations sont très proches les uns des autres, avec toujours une plus faible valeur pour le niveau individuel.

Les figures 2 et 3 montrent respectivement au niveau individuel et au

niveau agrégé les comparaisons entre les nombres d'émigrants prévus aux nombres observés dans chaque région. Si le modèle « micro » prédit exactement le nombre d'émigrants de la région Ile-de-France, il surestime for-

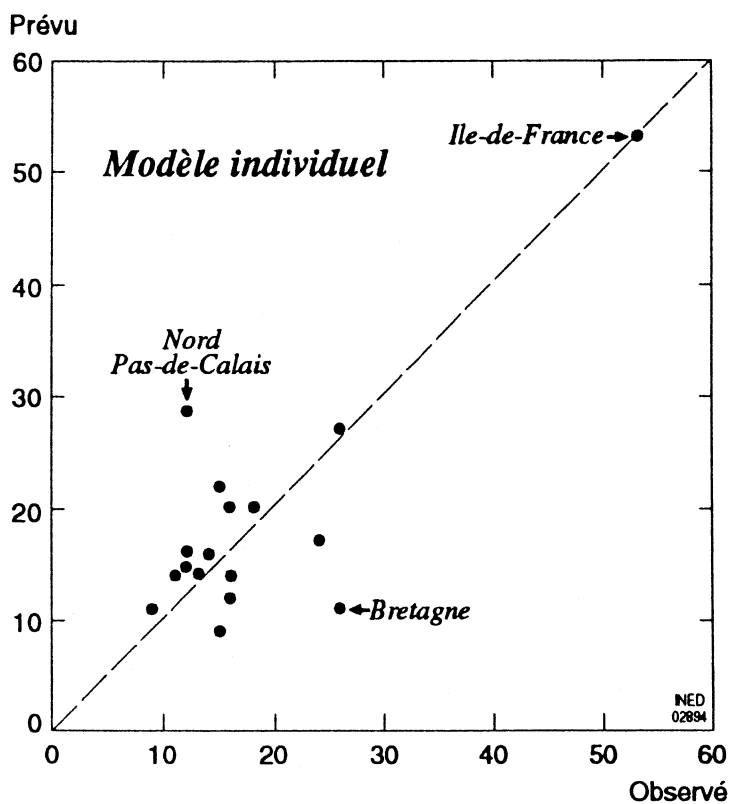


Figure 2 – Émigrants prévus par le modèle individuel comparés aux observés

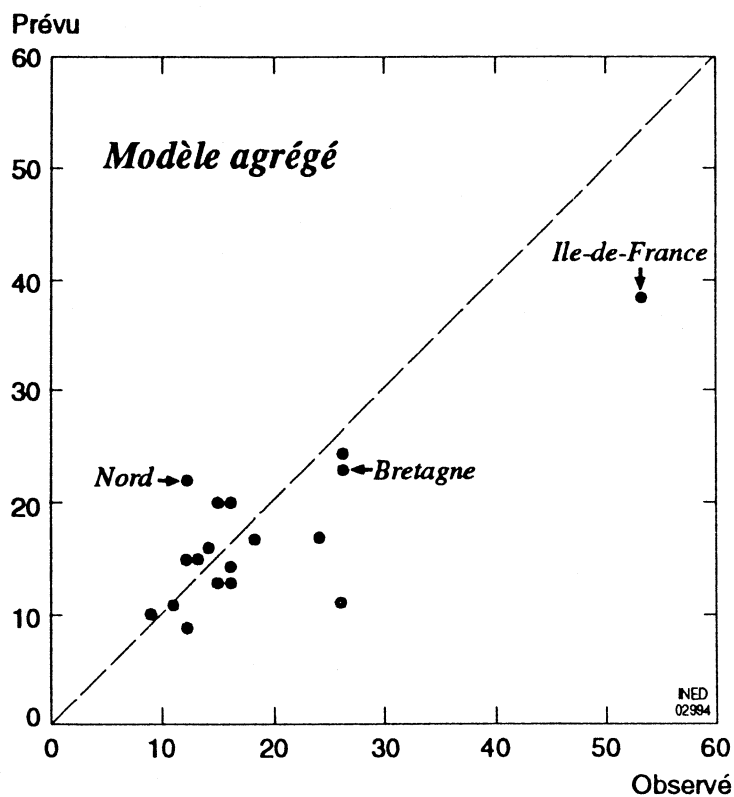


Figure 3. – Émigrants prévus par le modèle agrégé comparés aux observés

tement celui de la région Nord-Pas-de-Calais et sous-estime fortement celui de la Bretagne. Les émigrants des autres régions sont à peu près correctement prévus. Le modèle « macro » en revanche sous-estime fortement le nombre d'émigrants de l'Ile-de-France. Il surestime les émigrants de la Bretagne, comme le modèle « micro », les autres effectifs étant à peu près correctement prévus, un peu mieux même que dans le modèle « micro ».

Les deux types de modèles ne donnent donc pas des résultats absolument identiques lorsque l'on cherche à prévoir les effectifs de migrants.

Voyons maintenant les résultats que l'on obtient au niveau individuel, en faisant intervenir les caractéristiques de chaque enquêté. Le tableau 3 donne l'estimation des paramètres correspondants.

TABLEAU 3. — ESTIMATION DES PARAMÈTRES DU MODÈLE « MICRO » ET DE LEUR ÉCART TYPE, EN FAISANT INTERVENIR LES CARACTÉRISTIQUES INDIVIDUELLES

Caractéristique	Paramètre estimé	Écart-type
Mariée	- 0,568***	0,127
Propriétaire	- 1,106***	0,364
Inactive	0,339***	0,130
Agricultrice	- 0,188	0,181
Maximum du log. de vraisemblance	- 17 874,23	
*** Résultat significatif au seuil de 1 %		

On voit que trois caractéristiques ont un effet significatif sur la probabilité d'émigrer. Ces résultats sont cohérents avec les hypothèses que nous avons faites, même lorsque les paramètres estimés ne sont pas significativement différents de zéro : baisse confirmée de la probabilité de migrer pour les mariées et les propriétaires ; hausse confirmée de la probabilité de migrer des inactives ; pour les agricultrices, bien que le résultat ne soit pas significatif, baisse de la probabilité de migrer, comme pour les femmes logées chez leurs parents. Aucune relation linéaire n'apparaît entre les paramètres des modèles faisant intervenir les caractéristiques des zones et ceux faisant intervenir les caractéristiques individuelles. Il semblerait que ces deux types de caractéristiques jouent de façon très différente sur les probabilités d'émigrer.

Voyons si l'introduction simultanée des deux types de variables donne un modèle plus riche que le précédent. Le tableau 4 fournit ces résultats.

Ce modèle explique bien mieux la mobilité que le précédent : l'ajout des quatre caractéristiques agrégées entraîne une augmentation du maximum du logarithme de la vraisemblance de 12,56, correspondant à un χ^2 à quatre degrés de liberté égal à 25,12. En revanche, les paramètres estimés pour les deux séries de variables sont très proches de ceux obtenus lorsqu'interviennent séparément les caractéristiques agrégées (tableau 2) ou les caractéristiques individuelles (tableau 3). Cela vient confirmer l'hypothèse selon laquelle les deux types de caractéristiques joueraient de façon dif-

TABLEAU 4. – ESTIMATION DES PARAMÈTRES DU MODÈLE « MICRO » ET DE LEUR ÉCART-TYPE, EN FAISANT INTERVENIR SIMULTANÉMENT CARACTÉRISTIQUES INDIVIDUELLES ET CARACTÉRISTIQUES DES ZONES

Caractéristique	Paramètre estimé	Écart-type
Proportion de mariées	- 6,772***	1,918
Proportion de logées chez les parents	- 3,393***	1,430
Proportions d'inactives	5,383***	1,832
Proportion d'agricultrices	3,785***	0,802
Mariée	- 0,552***	0,128
Propriétaire	- 1,125***	0,364
Inactive	0,304**	0,131
Agricultrice	- 0,333*	0,185
Maximum du log. de vraisemblance	- 17 861,67	
*** Résultat significatif au seuil de 1 %		
** Résultat significatif au seuil de 5 %		
* Résultat significatif au seuil de 10 %		

férente sur les probabilités de migrer. Comme nous disposons des variances et covariances entre les divers paramètres estimés, il est possible de calculer les corrélations entre eux (tableau 5).

TABLEAU 5. – CORRÉLATIONS ENTRE CARACTÉRISTIQUES INDIVIDUELLES ET CARACTÉRISTIQUES DES ZONES DANS LESQUELLES VIVENT LES INDIVIDUS

Caractéristique	Corrélation
Mariée* Proportion de mariées	0,100
Logée chez les parents* Proportion de logées chez les parents	0,124
Propriétaire* Proportion de propriétaires	0,107
Inactive* Proportion d'inactives	0,100
Agricultrice* Proportion d'agricultrices	0,240

On voit que toutes ces corrélations sont faibles, en général de l'ordre de 0,10, indiquant une indépendance relativement forte entre ces caractéristiques « macro » et « micro ». La corrélation la plus élevée est celle estimée entre les paramètres mesurant l'effet d'être agricultrice et l'effet du pourcentage d'agricultrices sur la probabilité d'émigrer (0,24). Ce sont d'ailleurs ces paramètres qui changent le plus lors de l'introduction simultanée des deux catégories de variables. On constate, en particulier, que l'effet d'être agricultrice devient significatif au seuil de 10 %, et est opposé à celui du pourcentage d'agricultrices.

Ce paradoxe apparent peut s'expliquer en décomposant la population soumise au risque en deux groupes disjoints : les agricultrices et les non agricultrices. Il est dès lors possible d'estimer par régression le logarithme de la probabilité d'émigrer de ces deux groupes en fonction de la part d'agricultrices dans les diverses régions. La figure 4 porte ces résultats.

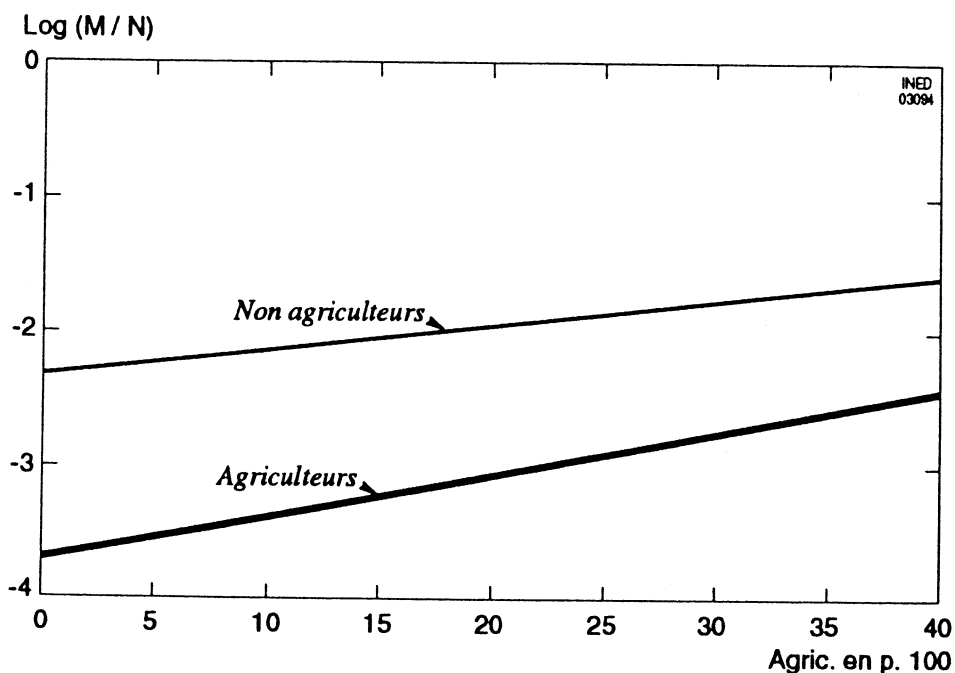


Figure 4. – Logarithme de la probabilité d'émigrer des agriculteurs et des autres catégories en fonction de la part d'agriculteurs dans chaque zone

On voit d'abord que les agricultrices ont toujours une plus faible probabilité d'émigrer, quelle que soit leur part dans la région. Cela confirme le paramètre négatif obtenu au niveau individuel ($-0,333$). Mais simultanément on voit que la probabilité d'émigrer, tant des agricultrices que des non agricultrices, croît lorsque la proportion d'agricultrices augmente. Il en résulte que le paramètre concernant cette variable agrégée est positif ($+3,785$). Le danger d'inférer certaines hypothèses sur le comportement individuel à partir de résultats obtenus au niveau agrégé, apparaît clairement : la présence de nombreuses agricultrices entraîne une plus forte probabilité d'émigration de toutes les catégories de la population, en partie à cause de la rareté des emplois non agricoles dans ces régions. Mais cela n'implique en rien que les agricultrices aient de plus fortes chances d'émigrer que les autres : c'est exactement l'inverse que l'on observe au niveau individuel. De plus quelle que soit la région d'origine, ce résultat est toujours vérifié.

Conclusions

La combinaison de résultats théoriques et de résultats concrets obtenus à l'aide de l'enquête « 3B », nous a permis de dégager un certain nombre de relations entre niveau individuel et niveau agrégé.

Tout d'abord il a été possible de relier les résultats obtenus aux deux niveaux lorsque les caractéristiques étaient de type agrégé. Ce résultat théorique a été confirmé par les données de l'enquête « 3B » : la corrélation entre les paramètres estimés aux deux niveaux est de 0,999 ; la pente de la droite de régression est cependant inférieure à l'unité. Cette différence doit provenir de ce que nous travaillons sur une longue période, l'égalité des paramètres n'étant vérifiée que si la période est très courte. On peut également se demander comment ces caractéristiques peuvent influencer sur les probabilités individuelles de migrer. A notre avis il s'agit d'une influence indirecte, et des enquêtes plus approfondies seront nécessaires pour montrer, par exemple, comment l'individu perçoit une forte proportion d'agriculteurs dans sa région de résidence.

Lorsque l'on fait intervenir les caractéristiques individuelles dans le modèle « micro », leur effet pourra être très différent de celui des caractéristiques mesurées au niveau agrégé. Nous avons montré, de façon théorique, qu'aucune relation simple n'apparaît entre ces paramètres et ce résultat a été confirmé empiriquement : la corrélation entre les paramètres estimés avec les caractéristiques des régions et ceux estimés avec les caractéristiques individuelles devient non significativement différente de zéro. Qui plus est, lorsque l'on fait intervenir simultanément les caractéristiques des régions et les caractéristiques individuelles, les divers paramètres restent inchangés, alors que la qualité du modèle s'améliore fortement. Nous avons pu vérifier que les corrélations entre caractéristiques des régions et caractéristiques individuelles sont très faibles.

Il semble, dès lors, que les deux types de caractéristiques aient un effet pratiquement indépendant sur les comportements individuels. Cela va à l'encontre de nombreux modèles qui étayaient leurs résultats sur des comportements individuels supposés, dont la somme sur la population conduit aux comportements et aux caractéristiques agrégés [Weidlich et Haag, 1988, pp. 11-20 ; Puig, 1981, pp. 49-50]. Comme on l'a vu ici, les zones où les proportions d'agriculteurs sont élevées ont une émigration bien plus forte que les autres, bien que les agriculteurs aient une probabilité beaucoup plus faible de migrer. L'effet de la caractéristique agrégée est dans ce cas opposé à la somme des comportements individuels.

Au vu de ces premiers résultats encourageants, il paraît nécessaire de poursuivre les recherches dans ce domaine. Nous allons indiquer ici les voies d'exploitation d'intérêt pour l'avenir.

En premier lieu, l'enquête « 3B » porte sur un trop faible effectif pour permettre une vérification correcte des résultats théoriques. L'utilisation de données de registres de population où sont recueillis des renseignements sur le mouvement de la population, devait améliorer cette vérification. Cependant, ces registres n'existent que dans un petit nombre de pays (la France n'en a pas) et ne prennent en compte qu'un petit nombre de renseignements individuels. Cela vient restreindre leur utilité.

En second lieu, nous avons dû faire ici de nombreuses hypothèses restrictives sur les modèles utilisés, de façon à pouvoir facilement les comparer. Il est nécessaire d'étendre nos résultats théoriques à un plus grand nombre de modèles. En particulier, il est important de faire intervenir les caractéristiques des zones de destination ainsi que celles mesurant l'interaction entre zone de départ et d'arrivée. Si ces caractéristiques sont relativement faciles à définir et mesurer au niveau « macro », il est beaucoup plus difficile de recueillir l'information que l'individu peut avoir sur les diverses destinations possibles, au niveau « micro ».

Comme on le voit les liens entre comportements individuels et comportements agrégés sont complexes et de nombreuses recherches sont encore nécessaires avant d'élucider complètement les relations entre ces deux niveaux. Les résultats obtenus ici montrent qu'il est cependant possible d'avancer dans ce domaine encore peu exploré.

Daniel COURGEAU

RÉFÉRENCES BIBLIOGRAPHIQUES

- COURGEAU, D., 1975.- « L'intensité des changements de catégorie de communes », *Population*, 30, n° 1, pp. 81-102.
- COURGEAU, D., 1985a.- « Interaction between spatial mobility, family and career life-cycle : A French Survey », *European Sociological Review*, vol. 1, n° 2, pp. 139-162.
- COURGEAU, D., 1985b.- « Changements de logement, changements de département et cycle de vie », *L'Espace géographique*, n° 4, pp. 289-306.
- COURGEAU, D., LELIEVRE, E., 1986.- « Nuptialité et agriculture », *Population*, 41, n° 2, pp. 303-326.
- COURGEAU, D. et LELIEVRE, E., 1989.- *Analyse démographique des biographies*, Editions de l'INED, Paris, 270 p.
- COX, D., 1972.- « Regression models and life tables », *Journal of the Royal Statistical Society*, B, 34, pp. 187-220.
- GREENWOOD, M., 1975.- « Research on internal migration in the United States : a survey », *Journal of Economic Literature*, XII, pp. 397-433.
- HAMERLE, A., 1989.- « Multiple-spell regression models duration data », *Applied Statistics*, 38, n°1, pp. 127-138.
- KALBFLEISCH, J. et PRENTICE, R., 1980.- *The statistical analysis of failure time data*, John Wiley and Sons, New-York, 322 p.
- LANCASTER, T., 1990.- *The econometric analysis of transition data*, Cambridge University Press, Cambridge.
- PUIG J.-P., 1981.- « La migration régionale de la population active », *Annales de l'I.N.S.E.E.*, n° 44, pp. 41-79.
- SANDEFUR, G. et SCOTT, W., 1981.- « A dynamic analysis of migration : an assessment of the effect of age, family and career variables », *Demography*, 18, n° 3, pp. 355-368.
- SANDERS, L., 1992.- *Système de villes et synergétique*, Anthropos, Paris, 274 p.
- STILLWELL, J., 1975.- « Models of interregional migration : a review », *Working paper*, Department of Geography, Leeds, 60 p.
- STOUFFER, S., 1940.- « Intervening opportunities : a theory relating mobility and distance », *American Sociological Review*, 5, pp. 845-867.
- TUMA, N.B. et HANNAN, M.T., 1984.- *Social dynamics. Models and methods*, Academic Press, Orlando, 578 p.
- WEIDLICH, W. and HAAG, G. eds., 1998.- *Interregional migration*, Springer-Verlag, Berlin, 388 p.
- YOUNG, E.-C., 1924.- « The movement of farm population », *Cornell Agriculture Experiment Station Bulletin*, 426.

COURGEAU (Daniel). – **Du groupe à l'individu : l'exemple des comportements migratoires**

Selon qu'il modélise les comportements de groupes ou ceux d'individus, le chercheur en sciences humaines va partir d'hypothèses, d'objectifs, de formulations et de caractéristiques très différents. Ainsi, lorsqu'il se situe au niveau agrégé il va utiliser les caractéristiques globales des groupes étudiés pour comprendre leurs comportements. A l'inverse lorsqu'il se situe au niveau individuel, il va travailler sur les divers éléments de la biographie de chaque personne. Cet article cherche à relier les résultats obtenus à ces deux niveaux d'agrégation tant d'un point de vue théorique que pratique. Il utilise les données d'une enquête biographique qu'il exploite en utilisant soit les caractéristiques individuelles, soit les caractéristiques agrégées, en vue d'étudier les comportements migratoires.

COURGEAU (Daniel). – **From the group to the individual: what can be learned from migratory behaviour**

The hypotheses, objectives, formulations and characteristics which social scientists use in their research differ greatly, depending on whether they are studying group or individual behavior. At the aggregate level, they tend to use the overall characteristics of the group to figure out its behavior. Conversely, at the individual level, they focus on various elements of each person's biography. This article attempts to link the result obtained at these two levels of aggregation both from a theoretical and a practical point of view. It uses data from a biographical survey which have been looked at on both from the individual and the aggregate point of view, to study migration behaviour.

COURGEAU (Daniel). – **Del grupo al individuo: el ejemplo de los comportamientos migratorios**

El investigador en ciencias sociales parte de hipótesis, objetivos, formulaciones y características muy distintas según modelice los comportamientos de grupos o de individuos. Si se sitúa a nivel agregado utilizará las características globales de los grupos estudiados para comprender su comportamiento. Por el contrario, si se sitúa a nivel individual trabajará sobre diversos elementos de la biografía de cada individuo. El presente artículo pretende poner en relación los resultados obtenidos a ambos niveles de agregación tanto desde un punto de vista teórico como desde un punto de vista práctico. Con tal propósito se utilizan los datos de una encuesta biográfica; éstos son explotados a nivel individual y a nivel agregado con el objetivo de estudiar los comportamientos migratorios.