

La durée de résidence en France: essai de mesure à l'aide de fichiers EDF

Courgeau D*, Nedellec V**, Empereur-Bissonnet P**

* Institut National d'Études Démographiques (INED)

** Service des Études Médicales d'EDF et de Gaz de France

Introduction

L'absence de registres de population en France, nuit à la connaissance des migrations internes qui se produisent chaque année. En effet, les données de recensements ne permettent que des mesures d'effectifs de migrants sur les périodes intercensitaires. Entre deux recensements on ne peut que prolonger les tendances mises en évidence antérieurement, alors que les flux de migration varient rapidement au cours du temps (Baccaïni *et al.*, 1993). Or l'EDF dispose d'une source quasiment exhaustive sur la mobilité des chefs de ménage, qui fournit des informations longitudinales, jusqu'ici non exploitées par les démographes. L'entreprise s'intéresse à la mobilité des ménages français à l'occasion de recherches visant à estimer les risques sanitaires liés à l'exposition aux polluants présents dans les sols d'anciens sites industriels (Empereur-Bissonnet, 1996). L'objectif de cette note est de donner une estimation des durées de résidence à partir de ces informations fournies par EDF

Données du système de facturation EDF

La représentativité du fichier de quittancement national est de 95 % des consommateurs individuels d'électricité. Environ 5 % des habitants sont alimentés par des régies indépendantes. Le taux de couverture EDF le plus faible est relevé dans les régions suivantes: Alsace (48 %), Poitou-Charentes (79 %), Picardie et Lorraine (86 %), Aquitaine (90 %).

Le tarif bleu, dit *de clientèle ordinaire*, exclut les administrations, les services publics et les entreprises. Il regroupe les clients domestiques, les clients agricoles, et les clients professionnels (tertiaires) exerçant leur activité à domicile et ne faisant pas la demande d'une facturation séparée. Pour la France entière, les données agrégées ne permettent pas la distinction entre les trois types de clientèle. Toutefois on dispose d'une extraction au centième permettant de les différencier: la durée moyenne des contrats, calculée en transversal au cours de la période 1971-1996, varie en fonction de la clientèle domestique (13,1 ans), agricole (24,2 ans) et professionnelle (10,0 ans). Avec les données agrégées, la moyenne nationale sur la même période est de 13,1 ans. Il n'y a donc pas de différence avec le résultat obtenu pour la clientèle domestique uniquement.

On dispose d'un état annuel exprimé en nombre de contrats actifs le dernier jour de l'année, classés par année de création. Par différence avec les chiffres de l'année précédente on obtient le nombre de contrats fermés dans l'année. Ces données servent de base aux calculs des quotients de déménagement annuel. Les informations ne sont connues que pour les contrats ouverts après 1973, date de la mise en place du système informatique. Par ailleurs il n'est pas possible de connaître pour chaque contrat le statut d'occupation (locataire ou propriétaire, résidence principale ou secondaire), ni le type d'habitat (date de construction ; zone urbaine, périphérique ou rurale), ni les caractéristiques socioprofessionnelles du contractant. De même nous ne disposons pas du nombre de personnes vivant avec le titulaire.

Estimation longitudinale de la durée de séjour

Les données de registres de population permettent de suivre parfaitement la mobilité des populations étudiées. Ainsi en Norvège de telles données permettent de dater au jour près toutes les migrations survenues dans la population (Ginsberg, 1979). Cela permet de travailler, non plus sur les seuls chefs de ménage, comme dans le fichier EDF, dont la définition n'est d'ailleurs valable qu'à un instant donné, mais sur l'ensemble des individus de la population. Les données dont nous disposons ici, moins parfaites que les données de registre, permettront cependant d'approcher ces résultats.

Méthode

Nous allons d'abord utiliser des méthodes non-paramétriques pour estimer les fonctions de séjour, selon la durée T écoulée depuis la précédente installation. T est donc une variable aléatoire, dont on va estimer la fonction de séjour à un instant donné t :

$$S(t) = P(T \geq t) \quad [1]$$

et le quotient instantané :

$$h(t) = \lim_{dt \rightarrow 0} \frac{P(T < t + dt / T \geq t)}{dt} \quad [2]$$

Nous ferons ici l'hypothèse que ces quotients restent constants au cours de chaque période annuelle. L'estimation classique nécessite des données plus complètes que celles dont nous disposons, elle est présentée afin de montrer les hypothèses que nous devons faire par la suite. Des diagrammes de Lexis (figure 1) illustrent ces questions : ils sont construits en prenant pour abscisse les années successives et pour ordonnée les durées de séjour.

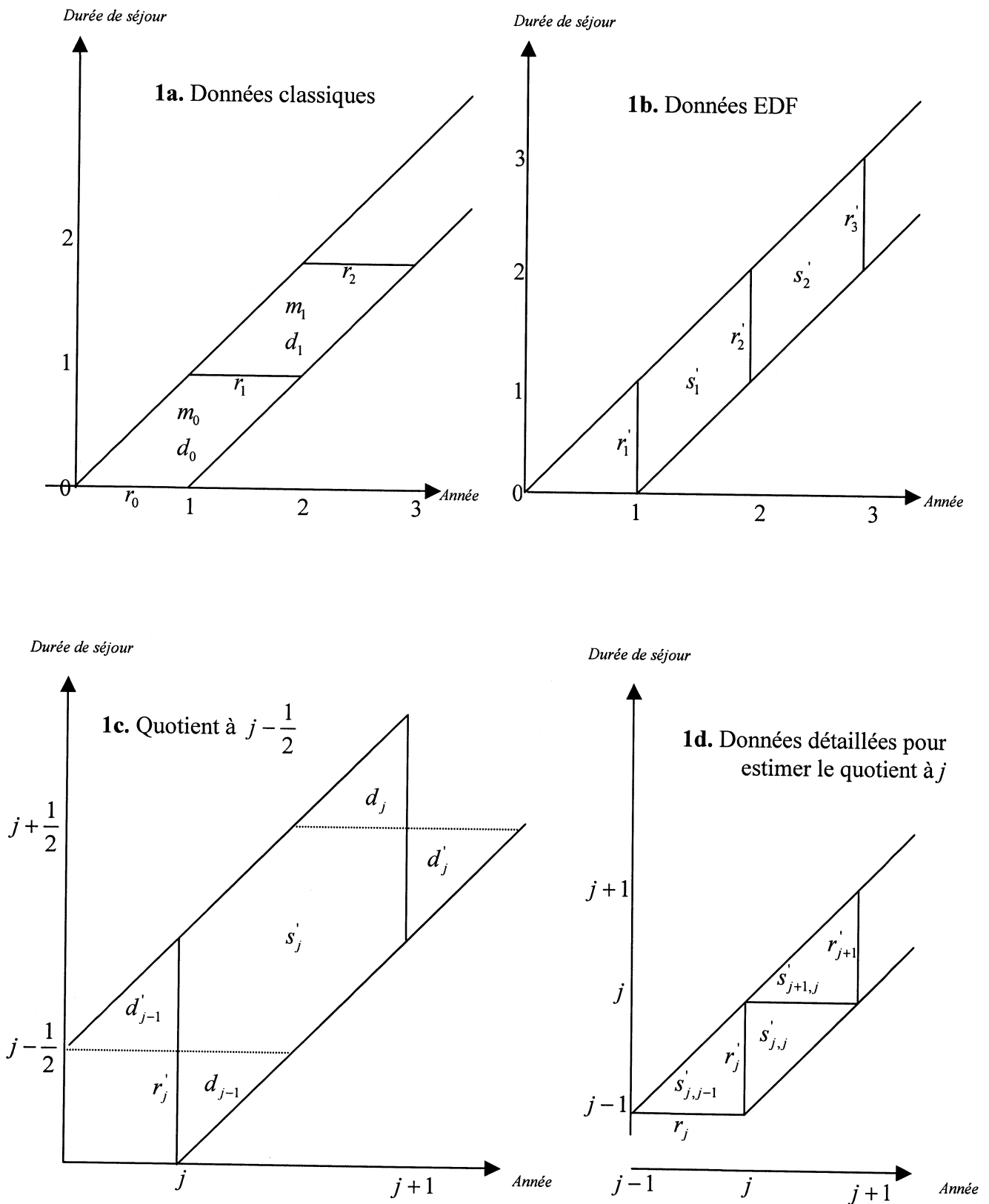


Figure 1 : Diagrammes de Lexis représentant les divers jeux de données pour estimer des quotients

Nous suivons une cohorte d'individus et observons au cours du temps la répartition de la migration suivante. La population soumise au risque au 1er janvier de l'année j à 0 heure est notée r_j . Au cours de l'année j les individus peuvent soit quitter leur logement (m_j), soit décéder (d_j). Ce type d'observation, parfait pour une étude longitudinale, est présenté sur la figure 1.a : dans chaque parallélogramme se trouvent les effectifs de décès et de migrations, en début se trouvent les populations soumises au risque. On montre que le quotient instantané, h_j , peut être estimé par la méthode du maximum de vraisemblance:

$$h_j = -\frac{m_j}{m_j + d_j} \log\left(1 - \frac{m_j + d_j}{r_j}\right) \quad [3]$$

On a également la possibilité d'estimer la fonction de séjour et le quotient cumulé, ainsi que les variances de tous ces estimateurs. Cependant les variances n'ont guère d'intérêt ici, étant donné que les observations sont quasiment exhaustives.

Les informations issues du fichier EDF se présentent différemment. Elles ne permettent pas de distinguer les sorties par nouvelle migration ou par décès des chefs de ménage. Dans ce cas on ne pourra estimer que des quotients de sortie sans distinguer la cause (migration ou décès).

$$h_j = -\log\left(1 - \frac{m_j + d_j}{r_j}\right) = -\log\left(1 - \frac{s_j}{r_j}\right) \quad [4]$$

De plus les nombres de départs sont enregistrés par année de survenue de l'événement et non par durée de séjour. Les départs de l'année initiale ne sont pas fournis : les individus ayant effectué une ou plusieurs migrations pendant cette année ne sont comptés qu'à partir de la dernière. La figure 1.b présente ces données. La population soumise au risque est maintenant fournie à la fin de chaque année d'observation. On voit qu'elles ne permettent pas d'estimer les quotients de sortie de la formule [4], sans hypothèses supplémentaires.

La figure 1.c présente la plus simple de ces hypothèses, qui conduit à travailler sur des intervalles différents des précédents. En effet, on peut supposer que l'effectif qui atteint la durée exacte $j - \frac{1}{2}$, égal à $r'_j + d'_{j-1} - d_{j-1}$, diffère peu de la population r'_j , lorsque d'_{j-1} et d_{j-1} sont approximativement égaux. Quant aux migrations et aux décès entre les durées exactes $j - \frac{1}{2}$ et $j + \frac{1}{2}$, ils sont à peu près égaux à ceux entre le 1er janvier et le 31 décembre, lorsque d_j est voisin de d'_j (Henry, 1972). Il en résulte une estimation approchée du quotient pendant cet intervalle:

$$\hat{h}_{j-\frac{1}{2}} = -\log\left(1 - \frac{s'_j}{r'_j}\right) \quad [5]$$

Si les courbes des quotients en fonction de l'âge sont suffisamment régulières, on a la possibilité d'estimer \hat{h}_j par interpolation linéaire entre $\hat{h}_{j-\frac{1}{2}}$ et $\hat{h}_{j+\frac{1}{2}}$. Par contre pour les fonctions de séjour, comme on ignore les événements qui se produisent dans le parallélogramme compris entre les durées 0 et 0,5 an, on devra faire des hypothèses

supplémentaires. La plus simple, utilisée ici, consiste à prendre l'instant initial à la durée 0,5 an. Une autre hypothèse serait de supposer que le quotient à appliquer pour ces durées est

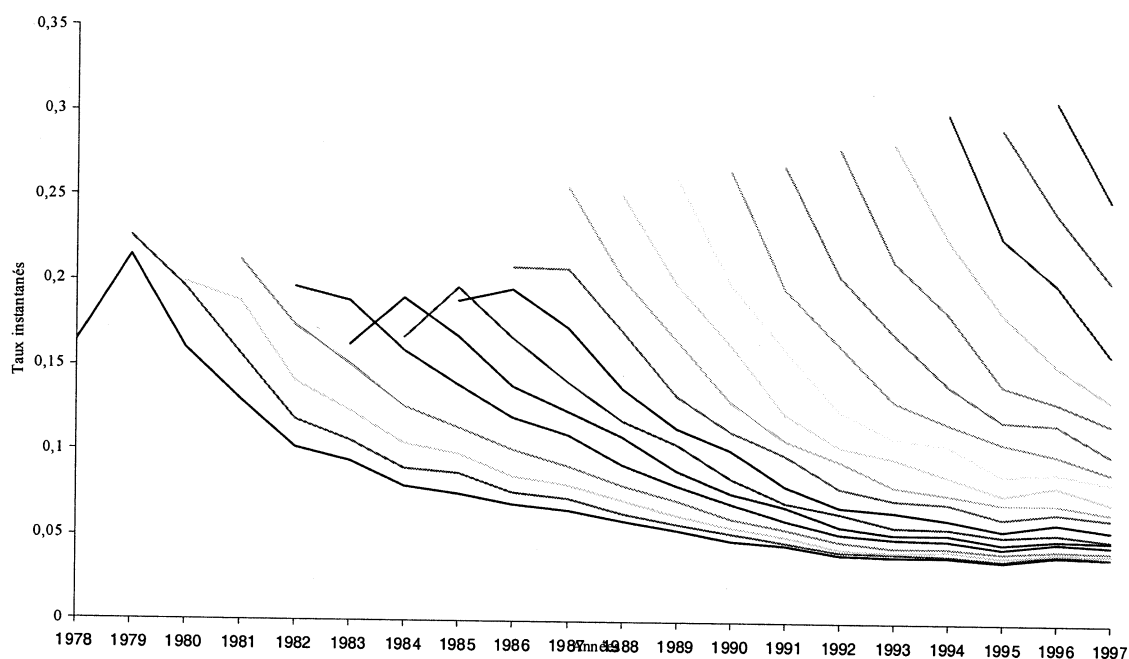
égal, par exemple, à $\frac{\hat{h}_{0,5}}{2}$ et d'estimer ainsi $\hat{S}_{0,5}^1$. Si l'on désire une estimation plus précise,

il faudrait disposer des données dans le découpage plus fin de la figure 1.d. Le calcul d'une vraisemblance correcte nécessite la connaissance de la population initiale et des événements situés dans le premier triangle de la figure 1.b : on les ignore ici. Nous n'irons donc pas plus avant dans cette estimation non-paramétrique et retiendrons la plus simple.

Résultats de l'estimation non-paramétrique

L'utilisation de la formule [5] nous permet d'estimer les quotients en fonction de la durée de séjour pour chacune des cohortes 1977-1996, portés sur la figure 2. Nous avons simultanément estimé les fonctions de séjour, en partant d'une durée initiale, en fait égale à 0,5 an, sur la figure 3.

Figure 2 : Quotients pour les cohortes installées de 1977 à 1995

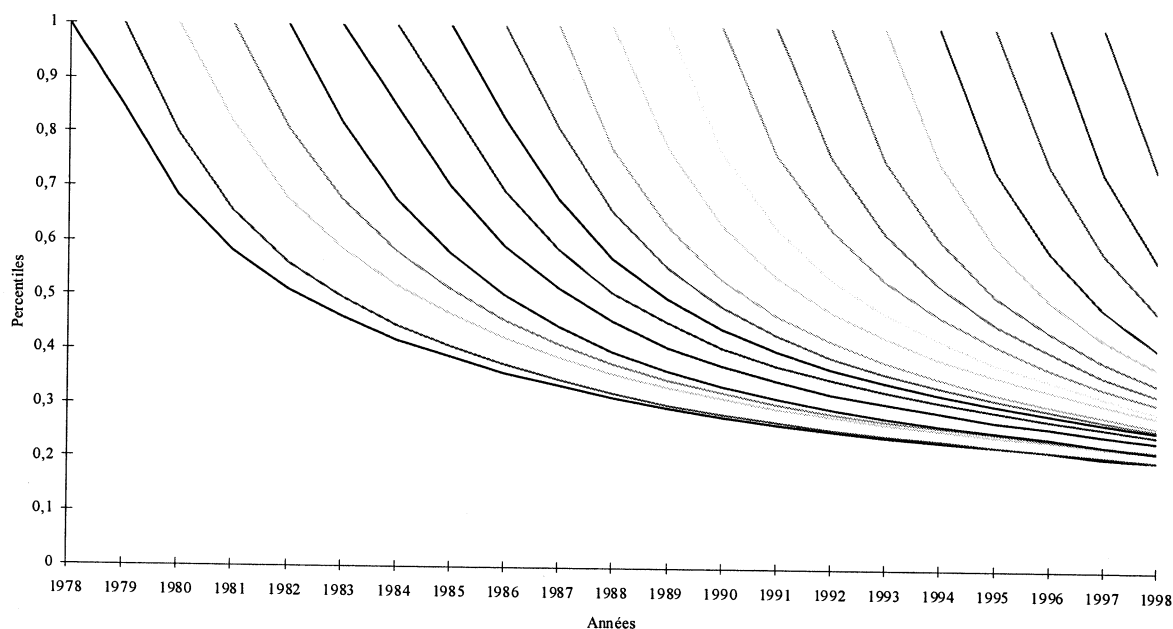


Les courbes donnant les quotients (fig 2) varient suffisamment pour démentir l'hypothèse d'un comportement identique des diverses cohortes. En premier lieu, on peut distinguer deux périodes : de 1977 à 1984, où les premiers quotients annuels tendent à diminuer d'une cohorte à l'autre, alors que dans la deuxième période, de 1985 à 1993, la tendance serait plutôt à la croissance. Les quotients initiaux fluctuent autour de 0,2 dans la première période et augmentent jusqu'à atteindre 0,3 lors de la cohorte de 1993 (deuxième période). Ensuite, des fluctuations annuelles importantes sont notées pour les courtes durées de résidence. Par exemple, pour les cohortes 1977, 1982, 1983 et 1984 le quotient instantané de la première année est inférieur à celui de la deuxième année. En revanche, les quotients de toutes les

¹ Voir Smith (1992), pour ces diverses méthodes d'estimation

cohortes étudiées décroissent de façon continue quand la durée augmente. Cette décroissance se ralentit, et les quotients semblent atteindre une valeur limite à peu près constante après une durée de 13 ans, proche de 0,05.

Figure 3 : Fonctions de séjour pour les cohortes installées de 1977 à 1995



L'examen des courbes donnant les fonctions de séjour (fig 3) vient compléter ces observations. On voit en premier lieu que la durée médiane de séjour est très courte, mais qu'elle passe d'une valeur proche de 4,25 ans pour les premières cohortes à une valeur inférieure à 3 ans pour les dernières. On voit combien cette durée est différente de la durée moyenne estimée en transversal, qui est autour de 10 ans en 1988 (Ministère de l'équipement et du logement, 1988). En fait, alors que l'on peut calculer une durée moyenne en transversal, puisque l'on dispose des durées de séjour de chaque migrant d'une année donnée, cette durée moyenne n'a plus aucun sens en longitudinal, car certains individus d'une cohorte donnée peuvent ne pas avoir encore migré à la date d'observation et pourront même ne jamais migrer après. On voit sur nos observations qu'après une durée de 20 ans, environ 20 % de la population n'a pas encore migré. Cependant, du fait que l'on mélange ici migrations et décès des chefs de ménage, la population d'une cohorte finira par s'éteindre du fait de la mortalité. Si en revanche on avait la possibilité d'éliminer cet effet de la mortalité, avec des données du type de celles de la figure 1.a, il est évident que la fonction de séjour pourrait ne jamais atteindre la valeur zéro. La durée médiane de séjour est toujours parfaitement définie.

Discussion

Ces résultats, bien qu'obtenus avec des données quasiment exhaustives, présentent un certain nombre d'incertitudes, si l'on veut décrire correctement l'évolution de la mobilité au cours de ces 20 dernières années. Voyons ici plus en détail les critiques que l'on peut faire aux résultats présentés dans la partie précédente.

En premier lieu, si la notion de chef de ménage a un sens précis au moment du déménagement

initial, celle-ci perd de sa pertinence au fur et à mesure que la durée de séjour va augmenter. Prenons par exemple un ménage constitué au départ d'un couple et de ses enfants. Si au cours du temps le couple se sépare mais que le chef de ménage reste seul dans son logement, le fichier EDF n'enregistrera aucun départ, alors que son ex-compagne et ses enfants ont fait une migration. Si au contraire le chef de ménage quitte ce logement pour le laisser à son ex-compagne et à ses enfants, le fichier EDF enregistrera alors sa migration. Le ménage initial, lui, n'existe plus après l'un ou l'autre de ces départs.

En second lieu, l'incertitude de ces données vient de ce qu'elles ne fournissent pas l'ensemble des effectifs nécessaires au calcul correct d'une fonction de séjour. La méconnaissance de l'effectif du premier triangle de la figure 1.b, et la mesure des effectifs de migrants par année de migration et non par durée de séjour, ne permettent qu'un calcul approché des quotients de migration sur des périodes inhabituelles, allant du milieu d'une année au milieu de la suivante. Des hypothèses sont nécessaires pour mieux ajuster ces résultats, mais la connaissance imparfaite des migrations de chefs de ménage, comparée à celle de la mortalité des populations humaines, rend ces hypothèses hasardeuses.

Nous allons cependant essayer de prévoir la migration des chefs de ménage à l'avenir. Nous baserons cette tentative sur des résultats graphiques sans tenter une recherche plus analytique d'un modèle prévisionnel.

Une prévision des migrations à venir

Divers essais graphiques nous ont montré que les modèles qui s'adaptent habituellement bien aux migrations individuelles (modèle de Gompertz, modèle migrants-sédentaires), étaient moins satisfaisants pour les données par ménage. Parmi les meilleurs ajustements empiriques, on constate qu'en portant la racine carrée de l'opposé du logarithme de la fonction de séjour en fonction du logarithme de la durée, on obtient des droites à peu près parallèles pour chaque cohorte. Cela revient à écrire la fonction de séjour sous la forme paramétrique suivante :

$$S(t) = e^{\left[-\beta^2(\log(\alpha) + \log(t))^2\right]} \quad [6]$$

L'estimation des paramètres α et β est faite manuellement sur les courbes elles-mêmes. La figure 4 porte d'une part la fonction de séjour estimée pour la cohorte 1977, que l'on peut entièrement comparer à la fonction observée, déjà donnée dans la figure 3, et celles estimées pour les cohortes 1986 et 1993, qui vont permettre de prévoir le comportement à venir de ces individus. On constate d'abord que l'ajustement est très correct. Ces courbes nous montrent dès lors une très forte baisse des pourcentages de non-migrants, que l'on peut prévoir au bout de 20 ans, qui passent de 19,94 % pour la première cohorte à 14,94 % pour la dernière. Les valeurs de $\log(\alpha)$ et de β pour les cohortes de 1977, 1986 et 1993 sont respectivement : 1,45 et 0,286 ; 1,57 et 0,285 ; 2 et 0,276.

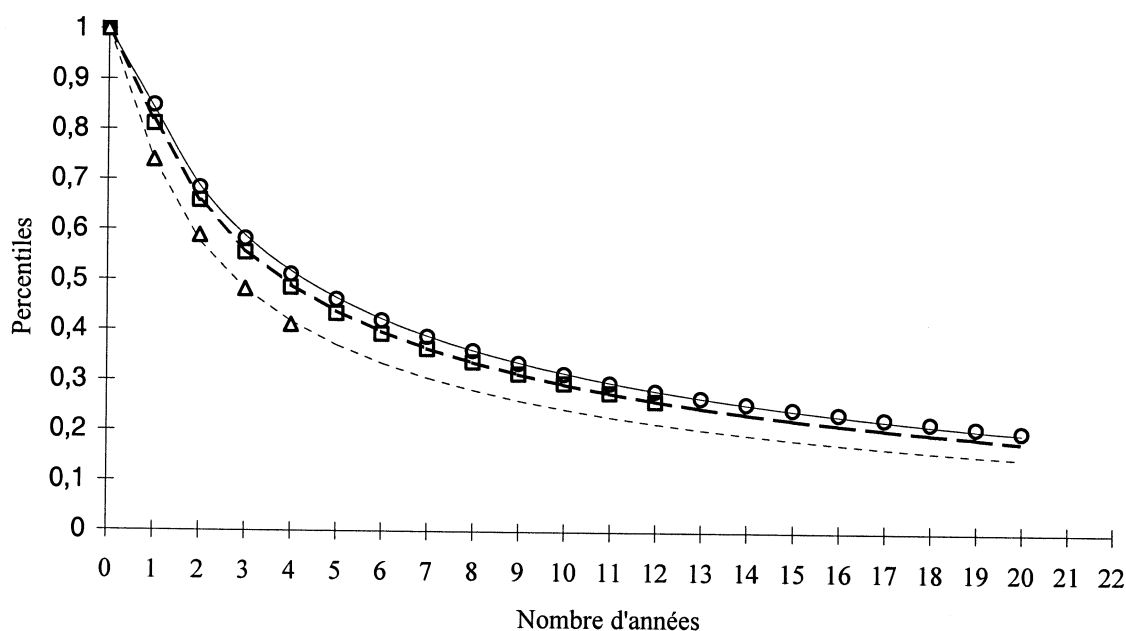


Figure 4. – Fonctions de séjour observées et estimées pour les cohortes installées en 1977, 1986, 1993

Conclusions

L'utilisation des données des fichiers EDF nous a permis, sous certaines hypothèses, d'estimer les fonctions de séjour des ménages français dans leur logement. Nous avons ainsi pu voir que cette mobilité a fortement augmenté pour les cohortes ayant fait une migration antérieure après 1986. On peut penser que cette augmentation apparaîtra² dans les résultats du prochain recensement en 1999, même si nous avons vu que les mesures effectuées avec les deux sources étaient très différentes. Cela constituera, de toutes façons, un test de la valeur des données EDF, pour mettre en évidence une évolution dans la mobilité des français sans avoir à attendre les résultats d'un recensement.

Nous avons également pu voir une assez grande régularité des courbes des fonctions de séjour, qui pourrait permettre des estimations du comportement futur de migrants. C'est à nouveau un résultat appréciable, pour juger en particulier de l'effet de politiques du logement, ou de l'effet de crises économiques, sur le comportement migratoire des ménages.

Ces données françaises sont aussi utilisables lors d'études prospectives d'évaluation des risques sanitaires. Le risque y est défini comme la probabilité de survenue d'effets néfastes pour la santé. Il est notamment influencé par la durée d'exposition aux facteurs délétères. Elle correspond à la durée de résidence lorsque l'exposition a lieu au domicile. Les résultats présentés ici permettront de réduire l'incertitude liée à l'usage de données françaises plus approximatives. Toutefois, dans les études spécifiques, afin de représenter correctement la situation locale, il faudrait pouvoir stratifier les cohortes selon des divisions géographiques du territoire national. La mise en place de statistiques plus détaillées, par département par

² Les données de l'enquête sur l'emploi montrent également une augmentation de la mobilité dès 1986 (XVIII^{ème} Rapport, 1989)

exemple, est possible, à un coût qui ne devrait pas être très élevé.

Remerciements

Cette étude a été rendue possible grâce au Projet de Développement de Service à la Clientèle Courante de la Direction Générale d'EDF GDF Services, qui a fourni les bilans annuels de facturation EDF. Nous les remercions pour leur collaboration constructive.

Références

Baccaïni B., Courgeau D., Desplanques G. (1993), “ Les migrations internes en France de 1982 à 1990. Comparaison avec les périodes antérieures ”, *Population*, n° 6, pp. 1771-1790.

Dix-huitième Rapport sur la Situation Démographique (1989), Paris : INED, 68 p.

Empereur-Bissonnet P. (1996), “ Usine à gaz de Genevilliers. Evaluation des risques sanitaires après réhabilitation du site. Rapport final d'étude ”, *EDF-GDF/SEM*, Paris.

Ginsberg R. (1979), “ Timing and duration effects in residence histories and other longitudinal data: II. Studies of duration effects in Norway, 1965-1971 ”, *Regional Sciences and Urban Economics*, pp.369-392

Henry L. (1972), *Démographie, analyse et modèles*, Paris, Larousse.

Ministère de l'équipement et du logement (1988), “ La durée moyenne d'occupation d'un logement est de 10 ans ”, *Notes vertes économiques*, n°20, pp. 1-4.

Smith D. (1992), *Formal demography*, New York, Plenum Press.