

Analyse des données biographiques erronées

In: Population, 46e année, n°1, 1991 pp. 89-104.

Citer ce document / Cite this document :

Courgeau Daniel. Analyse des données biographiques erronées. In: Population, 46e année, n°1, 1991 pp. 89-104.

http://www.persee.fr/web/revues/home/prescript/article/pop_0032-4663_1991_num_46_1_3669

Résumé

Courgeau Daniel. - Analyse des données biographiques erronées. A l'aide des données recueillies auprès de chaque époux, des conjoints réunis et des registres de population, qui portent sur divers événements de la vie familiale et migratoire des enquêtes, cet article met en évidence l'effet des erreurs de datation sur les résultats des diverses analyses de ces biographies. En dépit de la mauvaise fiabilité des données rétrospectives, en particulier sur les migrations, les résultats sont peu discernables selon la source utilisée et les variables prises en compte. Il semble que ces erreurs ne modifient pas ou changent peu l'ordre logique dans lequel les événements se produisent. Les erreurs de datation jouent donc le rôle d'un bruit de fond d'où l'on peut extraire une information cohérente quelle que soit la source utilisée.

Abstract

Courgeau Daniel. - An Analysis of Errors in Biographical Data. In this paper the author studies the effects of errors in the data on the outcome of different analyses, when data have been collected from each spouse separately, from the spouses jointly, and from population registers. The data relate to events that have occurred in the subjects' life histories and to migration. In spite of the poor reliability of retrospective data - particularly those relating to migration - the results obtained from different sources are virtually the same. The errors do not affect the logical order in which the events occurred. They are rather a form of background noise, but it remains possible to obtain coherent information from all types of source.

Resumen

Courgeau Daniel. - Análisis de datos biográficos erróneos. A partir de los datos obtenidos de cada esposo, de los cónyuges reunidos y de los registros de población, que contienen diversos acontecimientos de la vida familiar y migratoria de los encuestados, este artículo, pone en evidencia el efecto de los errores en las fechas declaradas, sobre los resultados de diversos análisis de esas biografías. A pesar de la poca fiabilidad de los datos retrospectivos, particularmente de las migraciones, los resultados son ligeramente perceptibles según la fuente utilizada y las variables consideradas. Parecería que esos errores no modifican o cambian muy poco el orden lógico de los acontecimientos en el cual se producen. Los errores en las fechas declaradas desempeñan entonces, un rol secundario allí donde se pueda obtener una información coherente, así se utilice una u otra fuente.

ANALYSE DE DONNÉES BIOGRAPHIQUES ERRONÉES

L'article précédent⁽¹⁾ a fait apparaître deux types de discordance entre diverses sources permettant de reconstituer la biographie familiale et résidentielle des individus : certains événements sont omis et ceux qui sont cités peuvent être mal datés. Les écarts les plus notables affectent les parcours migratoires. En particulier, les enquêtes rétrospectives auprès des intéressés permettent d'établir des itinéraires résidentiels plus complets que le registre belge de population qui ne saisit pas les déplacements à l'étranger ; mais en contrepartie la datation des mouvements est plus fiable dans le registre bien que la procédure conduise à retenir la date d'inscription au registre plutôt que la date de migration elle-même. Dans tous les cas, les renseignements obtenus par enquête auprès de l'épouse sont plus complets et plus précis que ceux fournis par le mari et la qualité est encore améliorée lorsque les deux conjoints se concertent. Au total donc, quatre sources dont la confrontation nous permet d'apprécier les erreurs auxquelles nous pouvons nous attendre en utilisant des données biographiques recueillies rétrospectivement. Il importe maintenant de voir comment ces erreurs vont influencer sur les résultats d'analyses plus complexes. En effet, de telles données d'enquêtes ne sont pas recueillies pour établir des statistiques simples, comme celles données dans l'article précédent, mais pour réaliser une étude plus détaillée des biographies.

Nous pensons ici aux méthodes d'analyse des biographies (Courgeau et Lelièvre, 1989) qui se sont mises en place au cours des dix dernières années. Elles visent à étudier les interactions qui apparaissent entre divers événements de la vie individuelle tout en démêlant l'hétérogénéité des populations sur lesquelles on travaille. On peut se demander si les erreurs de datation observées ne vont pas influencer fortement sur les résultats d'une telle analyse, au point de pouvoir les rendre totalement invalides.

Nous avons la possibilité, à l'aide des données de cette enquête, d'effectuer les mêmes analyses sur les différentes populations observées : hommes, femmes, conjoints réunis et données de registres de population.

(1) Michel Poulain, Benoît Riandey et Jean-Marie Firdion, dans le même numéro.

La comparaison des résultats obtenus sera facile car ces méthodes permettent, non seulement, d'estimer les effets des diverses caractéristiques sur les quotients de passage d'un état à un autre, mais également leurs variances et covariances. Des tests sont dès lors possibles pour voir si les diverses sources d'information fournissent des résultats différents entre eux ou non.

Nous allons utiliser ici les méthodes non-paramétriques, paramétriques et semi-paramétriques pour mettre en évidence l'effet des erreurs de datation sur les résultats qu'elles fournissent. Nous effectuerons d'abord une analyse non-paramétrique des durées de séjour dans les différents logements occupés par le ménage après le mariage. Cela nous permettra, simultanément, de montrer qu'un modèle de type Gompertz est bien adapté à ces données. Nous utiliserons ensuite ce résultat qui permet de prendre en compte de façon paramétrique cet effet de durée, et ferons intervenir diverses caractéristiques des individus en début de séjour. Nous serons alors à même de montrer en quoi les analyses effectuées diffèrent selon l'échantillon sur lequel on travaille.

Nous terminerons par une analyse semi-paramétrique des interactions entre la première de ces migrations et la première naissance légitime du couple. Nous ferons intervenir dans ce modèle diverses caractéristiques de la femme en vue, à nouveau, de déceler les différences qui peuvent apparaître entre les diverses sources.

Nous aurons ainsi testé dans des situations variées l'effet de l'utilisation de données biographiques erronées sur diverses méthodes d'analyse démographique des biographies.

I. – Analyse des migrations après le mariage

Nous observons ici toutes les migrations effectuées depuis le mariage et disposons des durées de séjour dans les divers logements occupés. Ces durées sont mesurées dans l'enquête en mois écoulés depuis la précédente migration. Nous avons cependant éliminé les durées de séjour de moins de six mois dans un logement, qui correspondent à des étapes provisoires que les individus ont sans doute plus souvent tendance à oublier. Il s'agit de toute façon d'un petit nombre de séjours de peu d'importance pour l'analyse des migrations.

Bien entendu, certaines de ces durées de séjour sont observées jusqu'à la date de l'enquête, sans qu'un déménagement soit survenu avant ce moment : on dit que l'intervalle avant la prochaine migration reste ouvert. Les données sont «tronquées à droite» par les sorties d'observation et on doit les utiliser en tant que telles dans l'estimation des quotients instantanés (voir Courgeau et Lelièvre, 1989, pp. 44-57).

Nous avons ici 1262 durées enregistrées pour les hommes, 1312 pour les femmes, 1316 pour les conjoints réunis et 1193 pour les registres. Le plus faible nombre donné par les registres est en partie dû au mauvais enregistrement des migrations internationales et au fait que tous les séjours n'ont pu être recueillis sur les registres. Du fait que nous ayons éliminé les durées de séjour de moins de 6 mois, ces chiffres sont légèrement différents de ceux donnés dans l'article précédent.

Pour réaliser l'estimation des quotients instantanés de migration, nous les avons supposés constants tout au long d'une année d'observation, leur changement n'intervenant que lorsque l'on passe d'une année à l'autre. Supposons que l'on observe au cours du $k^{\text{ième}}$ mois et la $j^{\text{ième}}$ année un nombre de migrations égal à m_k et des sorties d'observation de non migrants au nombre c_k . Si l'effectif P_j ne connaît pas de migrations et est toujours observé en fin d'année j , alors le quotient instantané de migration h_j est estimé par la formule suivante :

$$\hat{h}_j = \frac{\sum_{k=1}^{12} m_k}{P_j + \sum_{k=1}^{12} \frac{2k-1}{24} (m_k + c_k)} \quad (1)$$

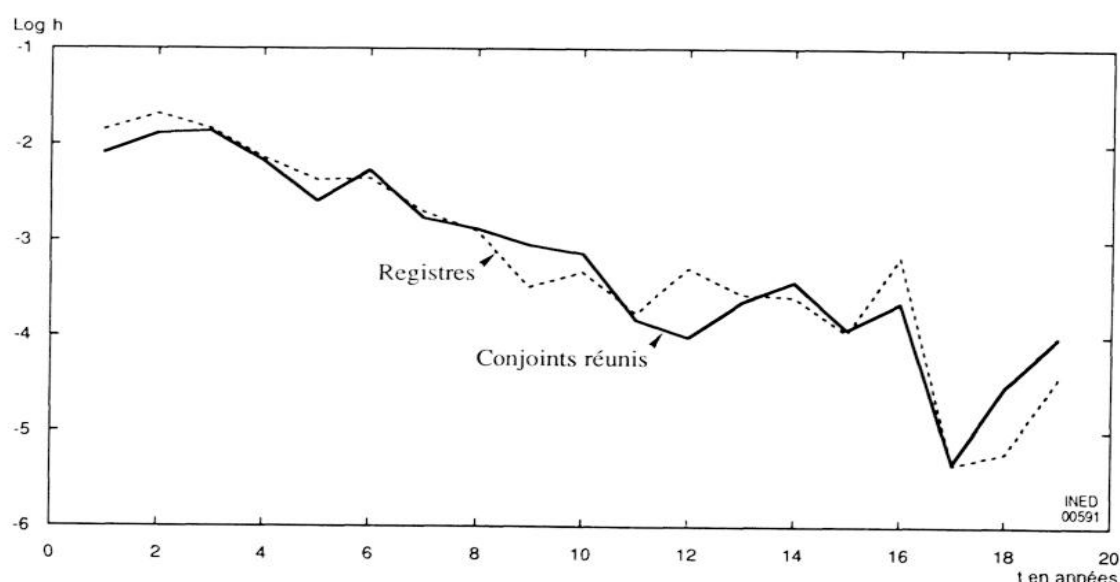
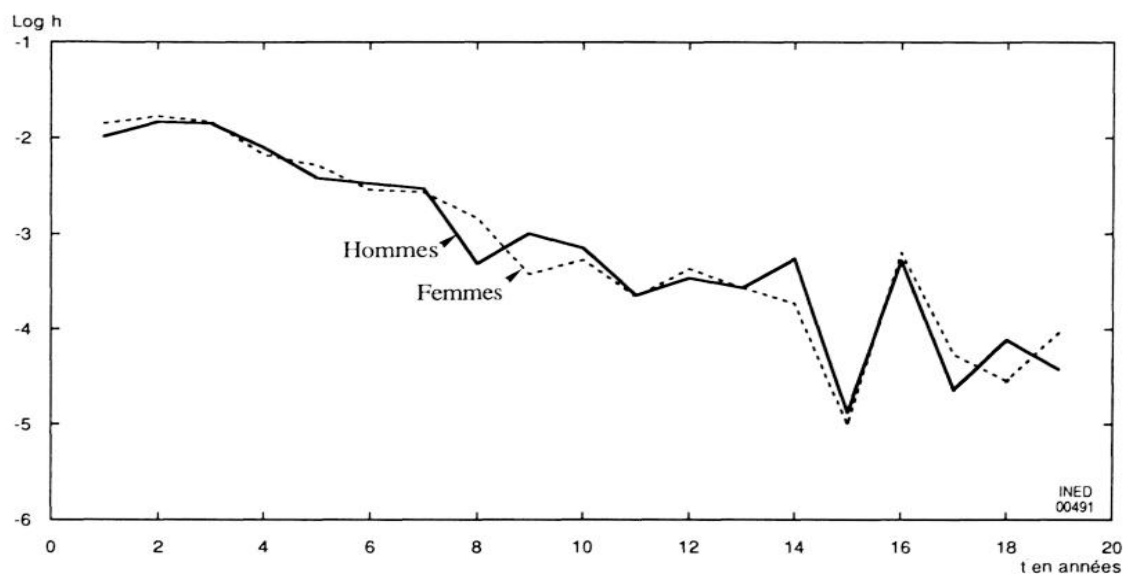
qui représente le nombre total de migrations observées l'année j divisé par le nombre total de personnes-années dans la population soumise au risque (Cox et Oakes, 1984, pp. 53-54).

Nous avons porté sur les graphiques 1 et 2, le logarithme népérien de ces quotients instantanés pour des durées s'étendant de un à 19 ans. On voit en premier lieu que ces courbes ne sont pas identiques du fait des erreurs dans le recueil des dates, selon qu'elles sont fournies par les hommes, les femmes, les conjoints réunis ou les registres. En dépit de ces différences, on observe que ces courbes semblent s'entremêler parfaitement : chaque courbe est tour à tour supérieure, inférieure aux autres ou comprise entre elles. On pourrait penser que les quatre sont issues de la même distribution sous-jacente et que les fluctuations que l'on observe sont dues aux faibles effectifs observés.

Nous avons la possibilité de tester cette hypothèse, en comparant les différences entre les migrations réellement observées dans chaque groupe et une quantité correspondant au nombre théorique de migrants que l'on observerait sous l'hypothèse d'identité des comportements (Courgeau et Lelièvre, 1989, pp. 65-66). Ce test peut être réalisé sur les 19 années d'observation prises simultanément et nous donne un chi-deux à trois degrés de liberté égal à :

$$\chi_3^2 = 1,045$$

qui ne contredit pas l'hypothèse précédente.



Egalement, l'examen de ces graphiques nous montre que l'on peut ajuster une droite sur ces données. Cela justifie l'utilisation d'un modèle de Gompertz pour l'analyse paramétrique qui suivra. On peut en effet écrire pour un tel modèle :

$$\log (h(t)) = \log (\lambda \rho) + \rho t \quad (2)$$

où ρ et λ sont des paramètres à estimer, t la durée de séjour considérée.

Les variations autour de ce modèle paraissent être plus importantes pour les durées élevées. Ces quotients ont en fait un intervalle de confiance beaucoup plus considérable que pour les faibles durées. Nous avons porté sur le tableau 1, le logarithme de ces quotients ainsi que l'écart-type de ce logarithme qui permet de vérifier que l'intervalle de confiance de ces logarithmes est supérieur aux variations observées sur les graphiques 1 et 2.

TABLEAU 1. – LOGARITHMES DES QUOTIENTS AINSI QUE L'ÉCART-TYPE
DE CES LOGARITHMES ESTIMÉS À PARTIR DE DONNÉES FOURNIES
PAR LES HOMMES, LES FEMMES, LES CONJOINTS RÉUNIS ET LES REGISTRES

Durée (années)	Hommes		Femmes		Conjoint réunis		Registres de population	
	log h	écart- type	log h	écart- type	log h	écart- type	log h	écart- type
1	-1,982	0,079	-1,844	0,073	-1,848	0,075	-2,093	0,086
2	-1,831	0,079	-1,772	0,077	-1,682	0,074	-1,887	0,082
3	-1,845	0,086	-1,828	0,084	-1,838	0,086	-1,860	0,089
4	-2,097	0,107	-2,174	0,114	-2,139	0,110	-2,172	0,116
5	-2,420	0,136	-2,285	0,124	-2,370	0,130	-2,593	0,146
6	-2,472	0,140	-2,541	0,146	-2,351	0,133	-2,268	0,128
7	-2,528	0,158	-2,565	0,156	-2,701	0,164	-2,767	0,172
8	-3,317	0,229	-2,840	0,186	-2,900	0,193	-2,881	0,193
9	-2,999	0,209	-3,422	0,243	-3,486	0,250	-3,044	0,218
10	-3,150	0,250	-3,275	0,243	-3,328	0,289	-3,168	0,267
11	-3,648	0,316	-3,659	0,374	-3,762	0,333	-3,827	0,354
12	-3,464	0,289	-3,368	0,277	-3,290	0,267	-4,015	0,378
13	-3,566	0,354	-3,566	0,376	-3,562	0,378	-3,639	0,378
14	-3,262	0,302	-3,733	0,353	-3,591	0,333	-3,433	0,333
15	-4,879	0,500	-4,998	0,490	-3,968	0,408	-3,928	0,447
16	-3,279	0,378	-3,197	0,353	-3,175	0,378	-3,652	0,408
17	-4,637	0,705	-4,270	0,564	-5,348	0,966	-5,328	0,975
18	-4,115	0,577	-4,543	0,674	-5,225	0,975	-4,526	0,706
19	-4,425	0,577	-4,036	0,571	-4,436	0,577	-4,015	0,577

Nous allons maintenant voir si un certain nombre de caractéristiques du ménage ont un effet sur cette durée de séjour et si la mesure de cet effet est influencée par le type de sources utilisées.

Pour cette analyse, nous faisons intervenir la durée entre le mariage et le début du séjour considéré (moins d'un an, de un à quatre ans, de cinq à neuf ans, le groupe témoin étant constitué par les séjours commencés dix ans et plus après le mariage), le nombre d'enfants nés avant le début du séjour. Nous n'avons pu faire intervenir le départ des enfants du domicile des parents étant donné le nombre trop faible de séjours concernés (moins de 8 séjours). Nous ferons, dans un second temps, intervenir le statut d'occupation du logement dans une analyse fondée sur les réponses des hommes, des femmes et des conjoints réunis (cette information n'est pas portée sur les données de registre), le groupe témoin étant dans ce cas composé essentiellement de locataires.

Nous utilisons pour ce faire un modèle paramétrique où la durée de séjour agit selon un modèle de Gompertz :

$$h(t; Z, \beta) = \exp(Z\beta + \rho t) \quad (3)$$

où Z est un vecteur ligne correspondant aux diverses caractéristiques de l'individu, β un vecteur colonne de paramètres à estimer, ρ représentant l'effet de la durée de séjour exprimée ici en années. La probabilité de migrer d'un groupe témoin était prise pour référence, l'effet d'un facteur est mesuré par l'exponentielle du paramètre β estimé pour ce facteur : ainsi lorsque ce paramètre est égal à + 0,485 pour les individus logés par leur employeur (tableau 2), cela signifie que ces individus ont une probabilité de migrer 1,62 (=exp (0,485)) fois plus forte que s'ils sont dans le groupe témoin, ici les locataires.

Le tableau 2 donne les résultats du premier modèle (caractéristiques du statut d'occupation exclues), estimés séparément sur les données fournies par les hommes, les femmes, les conjoints réunis et les registres de population⁽²⁾. Toutes les caractéristiques ont un effet très proche quelle que soit la source utilisée. Le nombre d'enfants en début de séjour n'a aucun effet sur la durée de séjour, comme nous l'avions observé sur la population française (Courgeau, 1985). En revanche, la durée de séjour influe fortement sur le quotient, venant confirmer les résultats obtenus en non-paramétrique, et la durée écoulée entre le mariage et le début du séjour joue également fortement sur ce quotient.

Nous avons porté sur le graphique 3 les coefficients estimés pour chaque caractéristique avec leur intervalle de confiance à 95 %. On y voit très clairement qu'en tenant compte de cet intervalle de confiance, il n'est pas possible de distinguer les résultats donnés par une source de ceux donnés par d'autres.

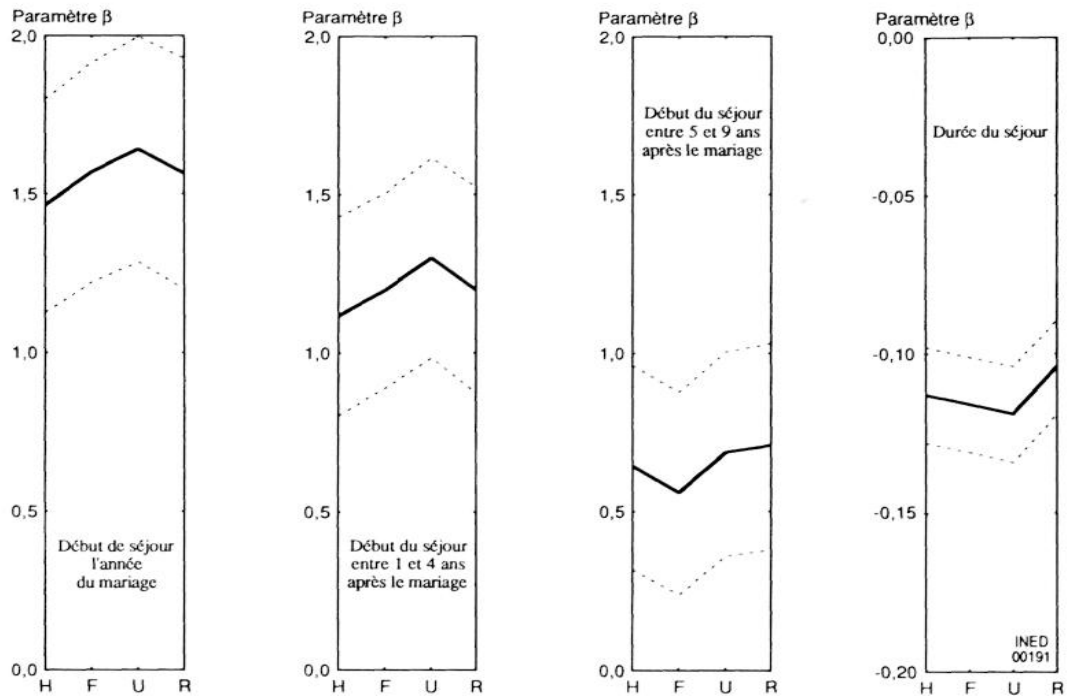
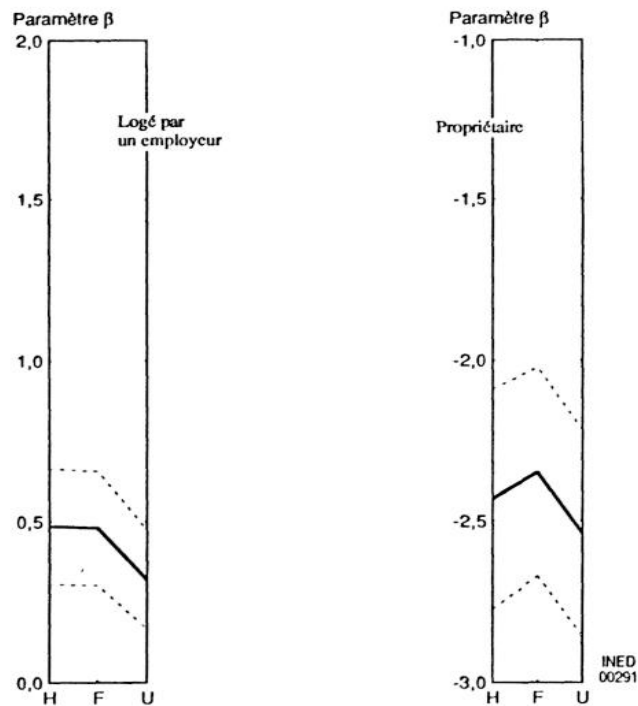
Ainsi, en dépit des erreurs sur les dates d'installation dans un logement puis d'émigration de ce logement, une analyse paramétrique conduit à des résultats très proches quelle que soit la source utilisée : chaque conjoint considéré séparément, les deux conjoints réunis et le registre de population.

Pour considérer d'autres caractéristiques, il nous faut laisser de côté les données de registre qui ne prennent que peu de renseignements sur les logements occupés. Or, il paraît important de tenir compte du statut d'occupation du logement qui doit avoir un effet important sur la probabilité de migrer d'un individu. Cette caractéristique, disponible seulement dans les enquêtes, permet une comparaison plus précise entre les probabilités de migrer. Nous avons porté, toujours dans le tableau 2, les résultats du

⁽²⁾ Les paramètres ont été estimés avec le programme RATE, mis au point par N. Tuma.

TABLEAU 2. – ANALYSE DE MOBILITÉ GÉOGRAPHIQUE : EFFET DE LA DURÉE ÉCOULÉE DEPUIS LE MARIAGE.
DE LA DURÉE DE SÉJOUR EN ANNÉES ET DU STATUT D'OCCUPATION SUR LA PROBABILITÉ DE CHANGER DE LOGEMENT
SELON LA SOURCE (PARAMÈTRES ESTIMÉS ET ÉCART-TYPE ENTRE PARENTHÈSES)

Variables	Hommes (1 260 séjours)		Femmes (1 310 séjours)		Conjoints réunis (1 314 séjours)		Registres (1 189 séjours)
	Modèle 1	Modèle 2	Modèle 1	Modèle 2	Modèle 1	Modèle 2	Modèle 1
Constante	- 2,955*** (0,174)	- 2,322*** (0,193)	- 3,009*** (0,179)	- 2,391*** (0,193)	- 3,062*** (0,183)	- 2,264*** (0,192)	- 3,118*** (0,187)
Début du séjour l'année du mariage	1,464*** (0,172)	0,438** (0,174)	1,569*** (0,177)	0,574*** (0,178)	1,642*** (0,182)	0,575*** (0,181)	1,564*** (0,186)
Début du séjour entre 1 et 4 ans après le mariage	1,117*** (0,160)	0,572*** (0,160)	1,198*** (0,157)	0,622*** (0,158)	1,301*** (0,161)	0,750*** (0,161)	1,199*** (0,166)
Début du séjour entre 5 et 9 ans après le mariage	0,641*** (0,164)	0,375** (0,165)	0,559*** (0,162)	0,286* (0,163)	0,684*** (0,164)	0,444*** (0,165)	0,707*** (0,166)
Nombre d'enfants en début de séjour	0,006 (0,040)	- 0,016 (0,039)	0,052 (0,042)	0,042 (0,041)	0,051 (0,042)	0,023 (0,042)	0,033 (0,044)
Logé par l'employeur		0,485*** (0,091)		0,480*** (0,090)		0,321*** (0,078)	
Propriétaire		- 2,431*** (0,175)		- 2,347*** (0,166)		- 2,538*** (0,164)	
Durée de séjour	- 0,113*** (0,0077)	- 0,056*** (0,0076)	- 0,116*** (0,0076)	- 0,058*** (0,0075)	- 0,119*** (0,0077)	- 0,064*** (0,0075)	- 0,104** (0,0076)
* Résultat significatif au seuil de 10 % ** Résultat significatif au seuil de 5 % *** Résultat significatif au seuil de 1 %							

Modèle 1*Modèle 2*

Graphique 3. – Paramètres estimés et intervalles de confiance à 95 % pour les diverses caractéristiques des modèles 1 et 2 du tableau 2 selon la source utilisée (H = Hommes, F = Femmes, U = Conjoints réunis, R = Registres)

second modèle qui utilise en plus cette information. Nous avons également porté sur le graphique 3 les coefficients estimés pour les individus logés par leur employeur et propriétaires, avec un intervalle de confiance de 95 %.

Ce nouveau modèle montre d'abord un effet tout à fait significatif du statut d'occupation du logement sur la durée de séjour, quelle que soit la source utilisée. Par rapport à la mobilité des locataires la probabilité de migrer est sensiblement plus forte lorsque l'on est logé chez son employeur et nettement réduite lorsque l'on devient propriétaire. Ces résultats sont tout à fait cohérents avec ceux obtenus avec l'enquête «triple biographie». Le graphique 3 montre, à nouveau, que les intervalles de confiance se recoupent entièrement.

L'effet des autres caractéristiques est réduit, mais reste semblable à ce qu'il était lorsqu'elles intervenaient seules : aucun effet du nombre d'enfants, augmentation de la probabilité de migrer lorsque le séjour débute moins de 10 ans après le mariage. A travers les réponses des femmes cette augmentation ne devient même significative entre cinq et neuf ans qu'au seuil de 10%. Cela est essentiellement dû au faible effectif enquêté, le signe du coefficient étant toujours positif. De toute façon, étant donné l'intervalle de confiance des coefficients, cet effet reste le même quelle que soit la source utilisée.

Des erreurs importantes sur les dates de migration et sur les durées de séjour n'entraînent donc pas de biais aussi importants sur l'analyse des probabilités de migrer en fonction de diverses caractéristiques de l'enquêté en début de séjour. Dans la majorité des cas, les résultats sont cohérents entre eux quelle que soit la source utilisée. Les quelques différences observées ne modifient pas les principaux résultats de l'analyse.

II. – Analyse des liens entre première naissance et première migration après le mariage

Nous allons maintenant considérer un cas plus complexe d'interaction entre deux phénomènes, l'un de la vie familiale (première naissance après le mariage), l'autre de la vie migratoire (première migration après le mariage). Nous combinons ainsi les erreurs sur les événements familiaux (mariage, naissance du premier enfant) et sur les événements migratoires.

Nous pouvons représenter cette étude par le schéma 1 bivarié.

Nous supposons ici, car nous ne disposons pas d'effectifs suffisants, que le quotient de fécondité de rang un ne dépend pas de l'âge à la première migration après le mariage ($h_{12}(t|u) = h_{12}(t)$) et que le quotient de première migration ne dépend pas de l'âge à la première naissance ($h_{21}(t|u) = h_{21}(t)$). Nous faisons également intervenir diverses caractéristiques, mesurées par

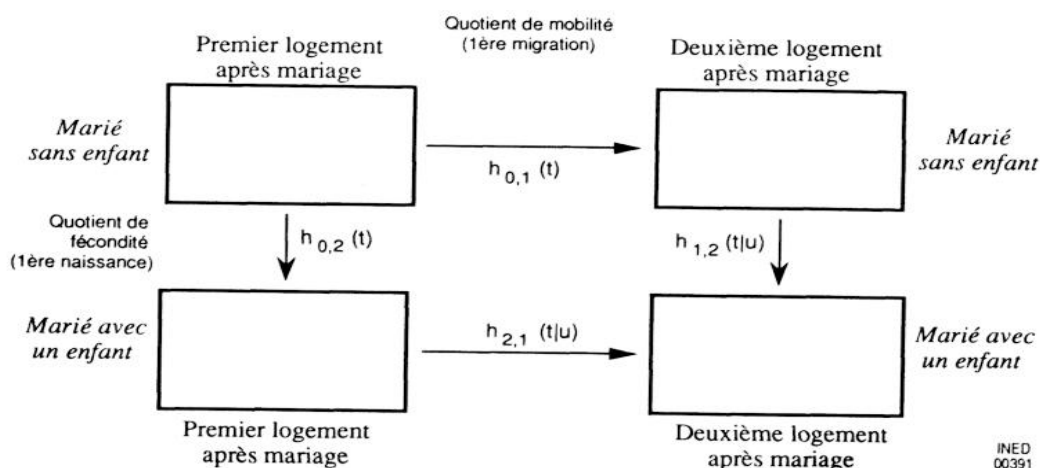


Schéma 1. – Première naissance et première migration après le mariage

un vecteur de variables, Z ou Z' , et nous utilisons directement un modèle semi-paramétrique, qui s'écrit :

$$\begin{aligned} h_1(t; \beta, Z) &= h_1(t) \exp(Z_1 \beta_1 + H(t-u)(\beta_0' + Z_1' \beta_1')) \\ h_2(t; \beta, Z) &= h_2(t) \exp(Z_2 \beta_2 + H(t-u)(\beta_0^2 + Z_2' \beta_2')) \end{aligned} \quad (4)$$

où le point représente les diverses situations possibles (0,1 ou 2), β_1 , β_1' , β_0 , β_0' , β_2 , β_2' , sont des paramètres à estimer pour les diverses caractéristiques avant (Z_1 et Z_2) ou après l'événement perturbateur (Z_1' Z_2'), $H(x)$ la fonction de Heaviside ayant la valeur 0, si x est négatif, ou la valeur 1, lorsque x devient nul ou positif. Les paramètres sont estimés par la méthode d'estimation de la vraisemblance partielle⁽³⁾, qui permettait ensuite d'estimer les quotients non-paramétriques $h_1(t)$ et $h_2(t)$.

Voyons d'abord l'effet de la première migration, considérée comme phénomène perturbateur, sur la naissance du premier enfant. Le modèle 1 du tableau 3 donne le paramètre estimé dans les différentes populations, lorsque la première migration intervient seule. Ils sont tous positifs et significativement différents de zéro, au moins au seuil de 10 %. Cela indique une augmentation de la probabilité d'avoir son premier enfant, une fois que l'on a migré. Cette augmentation est légèrement inférieure pour les données de registres, mais n'est pas significativement différente des résultats obtenus à partir des réponses des hommes, des femmes et des conjoints réunis.

Faisons maintenant intervenir les diverses caractéristiques de la famille, mesurées ici par les quatre sources. Nous avons distingué ici les couples selon l'âge au mariage de la femme, caractéristique que l'on sait influencer sur la fécondité : femmes mariées avant 20 ans, à 20-21 ans, à

⁽³⁾ Voir pour plus de détails sur ce modèle : Courgeau et Lelièvre, pp. 70-84 et pp. 155-165. Nous l'avons estimé à l'aide du programme EVACOV mis au point par E. Lelièvre.

22-23 ans et à 24-26 ans. Les femmes mariées après 26 ans constituent le groupe témoin. Les résultats de ce modèle 2 sont également portés sur le tableau 3.

Dans tous les cas, les paramètres estimés avant la migration sont positifs et toujours significativement différents de zéro. Cependant l'effet de l'âge au mariage de la femme est moins prononcé lorsqu'on utilise les informations données par leur mari. En particulier pour les femmes mariées avant 20 ans on a une fécondité de rang un beaucoup plus forte si l'on utilise l'information donnée par les femmes, les conjoints réunis et les registres, que lorsque l'on utilise l'information donnée par les hommes. Cela serait dû à une moins bonne mémorisation par les hommes de l'âge au mariage de leur femme.

Ce résultat est confirmé par le fait que, pour les hommes, la première migration n'introduit aucun changement dans ces effets, alors que pour les femmes, les conjoints réunis et les registres, cette première migration vient effacer significativement l'avantage des femmes mariées avant 20 ans : le coefficient d'interaction du tableau 3 est très faible pour les hommes et beaucoup plus significatif pour les autres sources.

Voyons maintenant l'effet de la fécondité sur la mobilité. Le tableau 4 fournit pour un premier modèle l'effet de la naissance du premier enfant lorsqu'elle intervient seule. Tous les paramètres sont positifs et significativement différents de zéro. Cela indique une augmentation de la probabilité de migrer après la naissance du premier enfant. Ce résultat est également cohérent avec ce que nous avons observé en France (Courgeau, 1985).

Faisons maintenant intervenir dans le modèle 2 (tableau 4) l'effet des diverses caractéristiques mesurées par les quatre sources, ici l'âge au mariage des femmes que nous avons déjà fait intervenir dans l'étude de la fécondité perturbée par la mobilité.

Ici à nouveau, les résultats obtenus pour les hommes diffèrent de ceux donnés par les autres sources. Pour les hommes, seul l'effet mis en évidence dans le modèle 1 reste significatif, l'âge au mariage de leur femme n'ayant pas d'effet significativement différent de zéro. Les autres sources sont cohérentes entre elles avant la naissance du premier enfant, plus forte probabilité de migrer des femmes mariées après 20 ans, qui va s'effacer après cette naissance ; pour les femmes mariées à 22-23 ans, plus faible mobilité après la naissance.

Ces résultats confirment la mauvaise mémoire des hommes sur l'âge de leur femme à leur mariage, qui vient effacer ici l'effet de cette caractéristique sur la première migration après le mariage.

TABLEAU 4. – EFFET DE LA NAISSANCE DU PREMIER ENFANT AINSI QUE DE DIVERSES CARACTÉRISTIQUES EN DÉBUT DE SÉJOUR
SUR LE QUOTIENT DE MOBILITÉ DE PREMIER RANG (PARAMÈTRES ESTIMÉS ET ÉCARTS-TYPES ENTRE PARENTHÈSES)

	Hommes (434 observations)			Femmes (443 observations)			Conjoints réunis (439 observations)			Registres (395 observations)		
	Effet principal β_1	Effet de la mig. β_0	Inter-action β_2	Effet principal β_1	Effet de la mig. β_0	Inter-action β_2	Effet principal β_1	Effet de la mig. β_0	Inter-action β_2	Effet principal β_1	Effet de la mig. β_0	Inter-action β_2
Modèle 1 : Naissance du premier enfant		0,311*** (0,128)			0,214** (0,129)			0,201* (0,131)			0,238** (0,143)	
Modèle 2 : Naissance du premier enfant		0,365** (0,181)			0,495*** (0,183)			0,393** (0,183)			0,443** (0,196)	
Femmes mariées avant 20 ans	- 0,034 (0,317)		- 0,067 (0,409)	0,530* (0,332)		- 0,764** (0,419)	0,429* (0,329)		- 0,695* (0,426)	0,600** (0,299)		- 0,918** (0,431)
Femmes mariées à 20-21 ans	0,000 (0,211)		- 0,178 (0,274)	0,151 (0,215)		- 0,431* (0,272)	0,026 (0,204)		- 0,250 (0,268)	- 0,009 (0,213)		- 0,160 (0,283)
Femmes mariées à 22-23 ans	- 0,135 (0,209)		- 0,020 (0,265)	0,190 (0,207)		- 0,465** (0,262)	0,099 (0,199)		- 0,334** (0,258)	0,194 (0,201)		- 0,431** (0,270)
Modèle 3 : Naissance du premier enfant		- 0,246 (0,241)			0,119 (0,248)			- 0,347* (0,230)				
Femmes mariées avant 20 ans	- 0,044 (0,317)		0,025 (0,410)	0,681** (0,335)		- 0,859** (0,423)	0,419 (0,331)		- 0,584* (0,430)			
Femmes mariées à 20-21 ans	0,019 (0,215)		- 0,168 (0,275)	0,241 (0,218)		- 0,433* (0,275)	0,036 (0,210)		- 0,099 (0,270)			
Femmes mariées à 22-23 ans	- 0,116 (0,210)		0,048 (0,267)	0,236 (0,208)		- 0,508** (0,263)	0,086 (0,199)		- 0,280 (0,259)			
Logée chez l'employeur	0,202 (0,174)		0,919*** (0,227)	0,490*** (0,182)		0,656*** (0,230)	- 0,074 (0,165)		1,213*** (0,217)			

* Résultat significatif au seuil de 10 %

** Résultat significatif au seuil de 5 %

*** Résultat significatif au seuil de 1 %

Faisons enfin intervenir le statut d'occupation du logement dans le modèle 3, toujours porté dans le tableau 4.

Le fait d'être logé chez l'employeur augmente la probabilité de migrer, surtout après la naissance du premier enfant, cela pour toutes les sources. L'effet de l'âge au mariage de la femme reste identique à ce que donnait le modèle 2.

Conclusion

L'article précédent montrait une mauvaise fiabilité des données rétrospectives, en particulier, sur les migrations, mais les premières analyses que nous venons d'effectuer modèrent ces conclusions négatives.

Les trois types d'analyse, non-paramétrique, paramétrique et semi-paramétrique, que nous avons faites ici permettent de nuancer ces résultats.

L'analyse non-paramétrique des durées de séjour dans les divers logements occupés depuis le mariage, nous a donné des résultats non discernables selon la source utilisée, comme l'a montré le test que nous avons effectué.

L'analyse paramétrique de ces mêmes durées de séjour en fonction de diverses caractéristiques de l'individu en début de séjour, conduit, également, à des résultats qui ne sont pas significativement différents, quelles que soient la source utilisée et les variables prises en compte. Les erreurs de recueil des diverses dates n'affectent en rien les résultats de cette analyse. La troisième analyse semi-paramétrique, portait sur l'interaction entre première naissance et première migration après le mariage.

Dans l'étude de la naissance du premier enfant, où la migration jouait le rôle de phénomène perturbateur, les résultats restent encore proches quelle que soit la source utilisée. On constate cependant un effet affaibli de l'âge au mariage des femmes, lorsqu'on utilise cette caractéristique telle qu'elle est donnée par les maris. Cela apparaît encore plus clairement dans l'analyse de la première migration après le mariage, où la naissance joue le rôle de phénomène perturbateur. La mauvaise qualité des réponses données par les hommes sur l'âge au mariage de leur femme rend cette variable non significative, alors que les trois autres sources continuent de donner des résultats cohérents entre eux.

On voit donc que c'est le recueil d'une des caractéristiques du conjoint, ici l'âge au mariage de la femme, qui donne les résultats les moins satisfaisants dans cette analyse semi-paramétrique des données fournies par le mari.

Il est dès lors utile de recommander le recueil des histoires de vie familiales et migratoires auprès des femmes, ou mieux des conjoints réunis lorsque cela est possible. La fiabilité des réponses est meilleure et conduit

à des résultats cohérents avec ceux donnés par le registre de population en Belgique. Bien entendu, pour l'histoire professionnelle, d'autres considérations peuvent jouer que nous ne pouvons pas tester ici car le registre de population ne saisit pas ces changements. L'utilisation d'autres sources est dans ce cas nécessaire (Bond et al., 1988). De toute façon, l'enregistrement de ces changements par les conjoints réunis devrait fournir la meilleure source.

Si les erreurs sur la datation des divers événements peuvent être importantes, il semble que ces erreurs ne modifient pas ou changent peu l'ordre logique dans lequel ces événements se produisent. Cet ordre est correctement remémoré, les erreurs de datation jouant le rôle d'un bruit de fond d'où l'on peut extraire une information cohérente quelle que soit la source utilisée. La mémoire semble donc fiable, là où l'analyse l'exige. Il est cependant préférable dans tous les cas, de recueillir l'information directement auprès de la personne qui a vécu l'événement, plutôt qu'auprès d'un tiers. Sous ces conditions, le type d'analyse auquel nous recourons peut être considéré comme satisfaisant. Cette conclusion ne préjuge évidemment pas de l'influence que pourraient avoir les erreurs de mémoire sur d'autres types d'analyse radicalement différents.

Daniel COURGEAU

BIBLIOGRAPHIE

- ANSIEAU D. (1988) – «Méthode longitudinale et méthode rétrospective à propos des questionnaires résidentiels. Fidélité des réponses rétrospectives», *Courrier des Statistiques*, 49 : 76-77.
- BOND G., BODNER K., SOBEL W., SHELLENBERGER R. and FLORES G. (1988) : «Validation of work histories obtained from interviews», *American Journal of Epidemiology*, 128, 2, 343-351.
- COURGEAU D., (1985, a) – «Interaction between spatial mobility, family and career life-cycle : a French survey», *European Sociological Review*, 1, 2 : 139-162.
- COURGEAU D. (1985, b) – «Effet de déclarations erronées sur une analyse de données migratoires», in *Migrations internes, collecte des données et méthodes d'analyse*, Chaire Quetelet 83, Jezierski ed. : 151-156.
- COURGEAU D. et LELIÈVRE E. (1986) – «Nuptialité et agriculture», *Population*, 41, 2 : 303-326.
- COURGEAU D. et LELIÈVRE E. (1989) – *Analyse démographique des biographies*, Manuel de l'INED.
- DUCHENE J. (1985) – «Un test de fiabilité des enquêtes rétrospectives - Biographies familiales, professionnelles et migratoires» in *Migrations internes, collecte des données et méthodes d'analyse*, Chaire Quetelet 83, Jezierski ed. : 135-150.
- HOEM J. (1985) – «Weighting, misclassification and other issues in the analysis of survey samples of life histories» in *Longitudinal analysis of labor market data*, Heckman and Singer eds., Cambridge University Press : 249-294.
- LYBERG I. (1983) – *The effect of sampling and non response on estimates of transition intensities : some empirical results from the 1981 Swedish Fertility Survey*, Stockholm research reports in demography, 14 : 1-75.

COURGEAU Daniel. – Analyse des données biographiques erronées.

A l'aide des données recueillies auprès de chaque époux, des conjoints réunis et des registres de population, qui portent sur divers événements de la vie familiale et migratoire des enquêtés, cet article met en évidence l'effet des erreurs de datation sur les résultats des diverses analyses de ces biographies. En dépit de la mauvaise fiabilité des données rétrospectives, en particulier sur les migrations, les résultats sont peu discernables selon la source utilisée et les variables prises en compte. Il semble que ces erreurs ne modifient pas ou changent peu l'ordre logique dans lequel les événements se produisent. Les erreurs de datation jouent donc le rôle d'un bruit de fond d'où l'on peut extraire une information cohérente quelle que soit la source utilisée.

COURGEAU Daniel. – An Analysis of Errors in Biographical Data.

In this paper the author studies the effects of errors in the data on the outcome of different analyses, when data have been collected from each spouse separately, from the spouses jointly, and from population registers. The data relate to events that have occurred in the subjects' life histories and to migration. In spite of the poor reliability of retrospective data – particularly those relating to migration – the results obtained from different sources are virtually the same. The errors do not affect the logical order in which the events occurred. They are rather a form of background noise, but it remains possible to obtain coherent information from all types of source.

COURGEAU Daniel. – Análisis de datos biográficos erróneos.

A partir de los datos obtenidos de cada esposo, de los cónyugues reunidos y de los registros de población, que contienen diversos acontecimientos de la vida familiar y migratoria de los encuestados, este artículo, pone en evidencia el efecto de los errores en las fechas declaradas, sobre los resultados de diversos análisis de esas biografías. A pesar de la poca fiabilidad de los datos retrospectivos, particularmente de las migraciones, los resultados son ligeramente perceptibles según la fuente utilizada y las variables consideradas. Parece que esos errores no modifican o cambian muy poco el orden lógico de los acontecimientos en el cual se producen. Los errores en las fechas declaradas desempeñan entonces, un rol secundario allí donde se pueda obtener una información coherente, así se utilice una u otra fuente.