



25 janvier 2017

Bernard AUBRY
Statisticien

Chiffres pour tous

Valoriser le nouveau recensement

Les enquêtes annuelles de recensement (EAR), diffusées en open data, offrent à tous l'opportunité d'effectuer des traitements de données qu'il n'aurait pas été envisageable d'effectuer il n'y a pas si longtemps. Cependant du fait du système de collecte, différent selon la taille des communes, et des choix faits dans le mode de diffusion (agrégation de cinq campagnes), les indicateurs établis sur les millésimes successifs des EAR n'ont pas de valeur conjoncturelle. L'objet de la communication est précisément de montrer que, sous certaines conditions, il est possible de pallier les inconvénients de cette méthode de diffusion et même d'obtenir une meilleure visibilité des phénomènes étudiés en dissociant dans les fichiers mis en ligne chacune des cinq collectes. On propose quatre exemples, donc trois sont directement liés à la connaissance des structures familiale. A chaque fois, les séries annuelles proposées portent sur 12 ans (2004-2015), au lieu de huit. Par ailleurs leur fiabilité semble établie, du moins si l'on en juge d'après la régularité des courbes. On aimerait pouvoir raccrocher ces résultats aux séries des recensements antérieurs (1962 à 1999), malheureusement les fichiers de ces collectes, pourtant harmonisés, ne sont pas proposés en libre accès.

L'Association internationale des démographes de langue française (AIDELF) a organisé, en liaison avec l'Institut de démographie de l'Unistra (Université de Strasbourg) son XIXème colloque les 21-24 juin 2016 sur le thème : « Configurations et dynamiques familiales ». Ce document mis en ligne sur le site de l'APR reprend la communication présentée lors de ce colloque et qui avait pour titre : « Apport et limites des EAR à la connaissance des structures familiales ».

Plan

I - La nécessité de séries longues et cohérentes

II - La diffusion en open data : une solution satisfaisante, mais pourquoi seulement depuis 2006 ?

III - L'intérêt d'isoler la variable campagne (année de collecte) : quatre applications

A - La mesure des taux d'activité dans un contexte de changement législatif

B - Les enfants d'origine étrangère : un demi-siècle d'observation (1968-2015)

**C - La taille de la famille, selon le statut de la femme, secteur public
ou privé (2004-2015)**

D - L'écart d'âge entre conjoints : les hommes nés en 1935 (2004-2015)

I - La nécessité de séries longues et cohérentes

Le contraste est frappant entre la richesse d'information acquise à travers un demi-siècle de recensements de la population dont les fichiers ont été informatisés et l'indisponibilité de séries chronologiques harmonisées qui auraient dû en être tirées et mises à disposition. Les canaux de distribution sont inadaptés aux besoins des utilisateurs. Même dans sa nouvelle version introduite en novembre 2016, le site de l'Insee n'est pas encore jugé satisfaisant par nombre d'utilisateurs, chercheurs ou non. Prenons un exemple, inspiré par la remarque faite par un professeur de démographie lors du colloque de Strasbourg. Jusqu'en 1999, l'Insee diffusait au lendemain de chaque recensement une série de documents papier (*fascicules bleus, orange, jaunes, verts...*) dont le contenu était stabilisé et donc bien intégré dans l'esprit des utilisateurs, notamment des enseignants. Rien n'aurait dû s'opposer à ce que le contenu des fascicules ne soit reproduit à l'identique sur le site, les chiffres ayant bien sûr été mis à jour. Car retrouver sur le site les chiffres des *fascicules verts* (données départementales) pour mettre à jour les séries sur lesquelles on a longtemps travaillé n'est pas une sinécure, même pour ceux qui disposent d'une certaine aisance dans l'art du surf sur internet. Quant à ceux qui n'ont pas d'expérience, ils sont privés d'une mise à jour de leurs séries patiemment composées !

Le contenu des fameux *fascicules orange*, qui fournissaient des séries communales sur des variables élémentaires telles que les populations, les logements, les ménages, l'état civil (naissances-décès) sont accessibles et mises à jour sur le site. Mais on aurait pu profiter des possibilités de l'informatique pour aller plus loin. On aurait pu fournir, depuis 1962 pour chaque commune, une gamme de données (par sexe, âge, niveau de formation), des données qui n'auraient aucun intérêt statistique pour les plus petites localités prises individuellement, mais qui permettraient d'effectuer des regroupements territoriaux spécifiques nécessaires à toute analyse des territoires.

Au fond ces propositions rejoignent le système de diffusion par menus déroulants. L'utilisateur compose lui-même le contenu des variables qu'il recherche : caractéristiques territoriales, structurelles et temporelles. Bien identifié dans l'ensemble du site, ce procédé est extrêmement convivial et donc accessible à tous.

II - La diffusion en open data : une solution satisfaisante, mais pourquoi seulement depuis 2006 ?

En revanche le site de l'Insee fournit un vaste ensemble de données en *open data*. L'utilisateur disposant de temps, d'un équipement adapté - aujourd'hui banalisé - et d'un minimum de savoir-faire est alors en mesure d'effectuer des traitements originaux et d'obtenir des résultats impensables il y a encore quelques années. Ces fichiers donnent à tous, et notamment aux chercheurs, la liberté de croiser à l'envi toutes les variables entre elles.

Cependant on ne comprend pas pourquoi l'Insee limite la mise en ligne de données individuelles aux seules enquêtes annuelles (EAR). Rien ne s'oppose à l'insertion des anciens recensements et notamment au fichier unique, harmonisé composé à partir des six collectes réalisées sur la période 1962 à 1999. Pour avoir accès à ces recensements, ils doivent effectuer d'inutiles démarches et procéder eux-mêmes à l'harmonisation des variables.

Les responsables du système statistique français devraient avoir à cœur de composer des fichiers cohérents non seulement pour les recensements mais aussi pour les différentes enquêtes périodiques, qu'elles relèvent ou non de la démographie (enquêtes sur l'emploi, sur le logement, sur la formation...). Bref, fournir des fichiers prêts à l'emploi. Car déchiffrer correctement les fichiers détail bruts, non harmonisés, est une perte de temps considérable pour l'utilisateur. Pourquoi les chercheurs devraient-ils refaire ce que d'autres ont fait avant eux ? Pourquoi mettre ces entraves à la diffusion des connaissances¹ ?

III - L'intérêt d'isoler la variable campagne (année de collecte) : quatre applications

Chaque enquête annuelle de recensement (EAR) est l'agrégation de 5 collectes annuelles, dont trois sont communes d'une année à l'autre. Les restrictions dans l'utilisation des fichiers énoncées par l'Insee se justifient par le système de collecte. En effet le territoire national est divisé en deux parties relevant chacune d'un mode de collecte tout à fait différent. La partie « *communes de plus de 10000 habitants* » est parfaitement aléatoire et la composition est indépendante d'une année sur l'autre. Ce n'est pas le cas de la partie formée par les localités plus petites : les données de ces communes, recalculées quand elles ne correspondent pas à l'année de collecte, ne sont pas indépendantes d'une année à l'autre. Cet argument plaide donc pour la plus grande prudence dans le traitement des fichiers mis en ligne.

Pourtant, il serait facile de créer et d'introduire dans les fichiers une variable « *année de collecte* », que pourrait appeler « *campagne* » et qui permettrait d'isoler dans un fichier EAR chacune de ses 5 composantes. Avec précaution, chaque collecte annuelle peut être vue comme un modèle réduit de la population du moment, bien que les fichiers annuels soient composés de deux ensembles obtenus différemment. Il n'est pas interdit de regarder en toute objectivité les indicateurs obtenus dans la

¹ Diffuser en open data des données individuelles pour les recensements depuis 2006, et restreindre pour tous, et notamment pour les chercheurs, l'accès aux recensements antérieurs n'est pas cohérent. Voir : <http://www.apr-strasbourg.org/membre/2054/aubry-bernard/document/717-diffusion-des-recensements-le-tmoignage-dun-chercheur.html>

série des campagnes annuelles de 2004 à 2015. Nous l'avons fait pour quelques variables démographiques et l'on constate que les séries annuelles sont parfois en mesure de mieux cerner les évolutions temporelles que les séries des EAR qui écrasent largement les effets conjoncturels. Pour convaincre de l'intérêt d'utiliser les campagnes plutôt que les EAR, nous proposons quatre exemples à savoir 1) le suivi des taux d'activité à 61 ans dans un contexte d'allongement de l'âge légal de départ à la retraite 2) l'évolution de la proportion d'enfants d'origine étrangère, 3) l'évolution du nombre d'enfants dans les familles et 4) l'évolution de l'écart d'âge entre les membres d'un couple.

A - La mesure des taux d'activité dans un contexte de changement législatif

Pour un âge légal inchangé (60 ans) et pour une génération donnée, les taux d'activité chutent assez brutalement de 80 % à 20%². L'impact statistique de ce changement rapide n'est pas correctement mesuré par les EAR car celles-ci écrasent les données de cinq générations aux comportements très différents. Quand le gouvernement décide de repousser de deux ans l'âge légal de départ à la retraite, et ce d'une façon progressive, l'évolution des taux d'activité produite par les EAR devient totalement indéchiffrable. En revanche, en dissociant chaque EAR en chacune de ses composantes annuelles, chacune fournissant une image représentative de la population du pays, on peut suivre dans le temps les effets de la mesure. L'étude porte sur les 8 enquêtes annuelles actuellement disponibles (2006 à 2013), ce qui fournit un jeu de 40 observations.

L'observation

L'indicateur utilisé est le taux d'activité par âge, ou par génération, rapport entre le nombre d'actifs et la population de référence. On s'intéresse aux personnes âgées de 61 ans (code AGED='061'). A la mi-janvier, au moment de la collecte, il s'agit pour l'essentiel de personnes nées en 1950 (collecte 2011) et qui sont a priori le plus touchées par les conséquences de la nouvelle loi.³ Le tableau 1 distribue les 40 taux d'activité selon leur EAR et leur campagne.

Tableau 1 – taux d'activité à 61 ans selon les EAR et les campagnes

Campagne	Génér.	millésime de l'EAR							
		2006	2007	2008	2009	2010	2011	2012	2013
2004	1943	20,2							
2005	1944	21,3	21,3						
2006	1945	22,3	22,3	22,2					
2007	1946	22,7	22,7	22,6	22,6				
2008	1947	23,4	23,3	23,3	23,3	23,2			
2009	1948		24,5	24,5	24,5	24,4	24,4		
2010	1949			26,9	26,9	26,9	26,8	26,8	
2011	1950				28,3	28,3	28,2	28,2	28,2
2012	1951					33,6	33,6	33,6	33,6
2013	1952						42,0	42,1	42,1
2014	1953							47,7	47,7
2015	1954								49,5

² Ainsi en 2006, les taux d'activité étaient les suivants : 74% à 56 ans, 42% à 60 ans, 11% à 64 ans.

³ Du moins pour la grande majorité d'entre elles, puisque la date de référence de la collecte se situe autour du 20 janvier.

taux d'activité	22,0	22,8	23,9	25,1	27,3	31,0	35,7	40,2
------------------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------

D'après fichiers détail EAR-2006-2013

Lecture : en ligne, les campagnes (années de la collecte), en colonne les EAR

La première diffusion des EAR est millésimée 2006 : les cinq campagnes donnent des taux d'activité en légère croissance (lecture en colonne, de 20,2 à 23,4%), une augmentation probablement à mettre en lien avec les préconisations du traité de Lisbonne (2000) qui promouvait une augmentation des taux d'emploi des seniors (55-64 ans). A l'EAR 2007, la campagne 2004 disparaît. Une nouvelle campagne, celle de 2009, entre à sa place dans l'EAR. Et ainsi de suite. La ligne du bas, moyenne des cinq campagnes donne le taux d'activité tel que diffusé sur le site de l'Insee.

Il ressort que les taux d'activité pour une même campagne sont quasiment stables d'une collecte à l'autre. En revanche, les chiffres en colonne augmentent très vite à partir du moment où la loi produit ses effets.

Les extrapolations

La stabilité des taux d'activité par campagne rend possible la prolongation de la série en ligne jusqu'à ce que l'on dispose de cinq valeurs, puisqu'une même campagne intervient cinq fois.⁴ La moyenne en ligne donne alors une estimation du taux d'activité annuel.

Tableau 2 - les taux d'activité observés et extrapolés

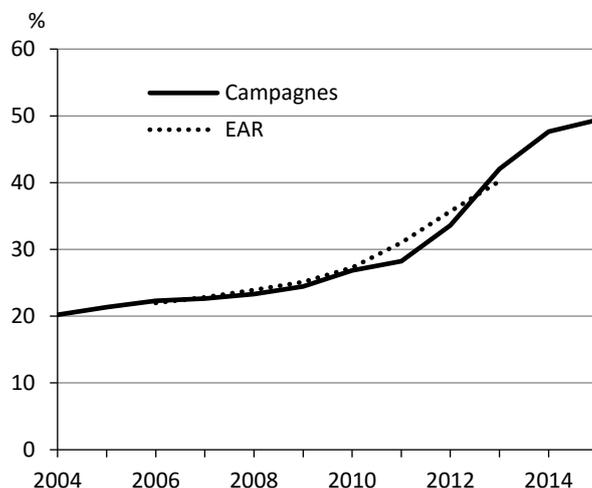
ANNEE	GEN.	ESTIM.	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
2004	1943	20,2	20,2	20,2	20,2	20,2	20,2											
2005	1944	21,4		21,4	21,4	21,4	21,3	21,3										
2006	1945	22,3			22,4	22,3	22,3	22,2										
2007	1946	22,7				22,7	22,7	22,7	22,6	22,6								
2008	1947	23,3					23,4	23,3	23,3	23,3	23,2							
2009	1948	24,5						24,5	24,5	24,5	24,4	24,4						
2010	1949	26,8							26,9	26,9	26,9	26,8	26,8					
2011	1950	28,2								28,3	28,3	28,2	28,2	28,2				
2012	1951	33,6									33,6	33,6	33,6	33,6	33,6			
2013	1952	42,1										42,0	42,1	42,1	42,1	42,0		
2014	1953	47,6											47,7	47,7	47,6	47,6	47,6	
2015	1954	49,4												49,5	49,4	49,4	49,3	49,3

D'après fichiers détail EAR 2006-2013

On dispose maintenant de deux séries de taux d'activité : une série de huit issues des EAR (tableau 1, ligne du bas) et une autre de douze qui sont le résultat d'observations et d'extrapolations (tableau 2, en colonne). Le graphique 1 compare l'évolution des taux d'activité selon le mode de calcul.

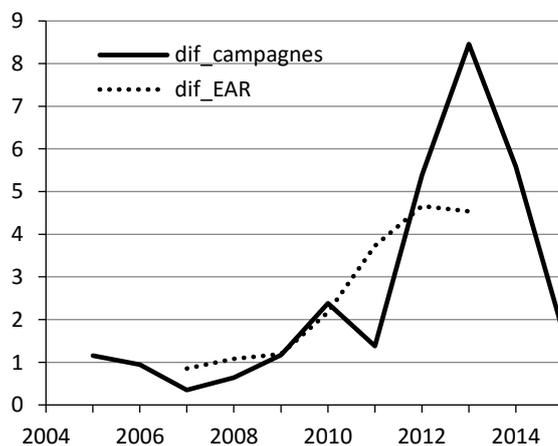
⁴ Convention : en ligne on calcule le rapport entre deux valeurs successives (quatre résultats). La moyenne en colonne de ces rapports sert au calcul des valeurs extrapolées. Noter que les coefficients sont très proches de 1 (de l'ordre de 0,996).

Graphique 1 - Comparaison des taux d'activité selon le mode de calcul (61 ans)



La courbe des EAR fait ressortir effectivement une tendance générale à l'allongement de la durée de la vie active, mais elle présente deux inconvénients : d'une part, elle gomme la conjoncture et d'autre part, elle couvre une période plus courte. A l'inverse, la courbe des campagnes donne une image de l'impact année par année de la loi à l'horizon 2015.

Graphique 2 – les variations annuelles, en points, des taux d'activité à 61 ans selon le mode de calcul



D'après fichiers détail EAR -2006-2013

Lecture : 2011, l'augmentation du taux d'activité par rapport à l'année précédente est de 1,7 point d'après les collectes annuelles et de 3,7 d'après les EAR.

Le graphique 2 compare les variations annuelles et, par conséquent, il accentue les contrastes entre les deux courbes. Dès la campagne 2012, l'accroissement du taux d'activité est forte, de l'ordre de 5 points, l'année suivante l'augmentation est plus forte encore (8 points). Les taux augmentent, quoique plus faiblement, lors des deux campagnes suivantes. Il faut voir là sans doute la conséquence de la progressivité de la loi. Une interrogation : pourquoi, la courbe « campagne » fléchit-elle en 2011 ? Est-

ce une variation aléatoire (très peu probable) ou faut-il chercher une explication économique ou comportementale, à la veille d'un changement de régime des retraites ?

B - Les enfants d'origine étrangère : un demi-siècle d'observation (1968-2015)

Contrairement au cas précédent, nous sommes dans une situation de régime stable, ce que confirme le tableau 3 qui présente la proportion de jeunes d'origine étrangère (0-17 ans).

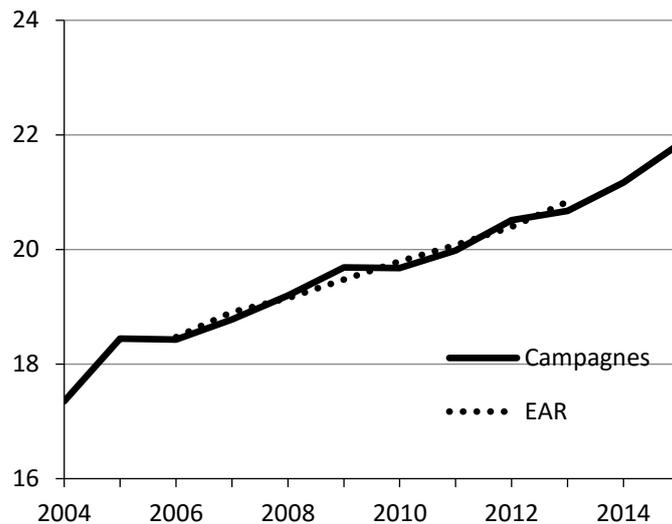
Tableau 3 - Proportion de jeunes d'origine (0-17 ans) d'origine étrangère par EAR et campagne – 2006-2013

ANNEE	2006	2007	2008	2009	2010	2011	2012	2013
2004	17,27							
2005	18,41	18,32						
2006	18,45	18,36	18,31					
2007	18,86	18,76	18,70	18,69				
2008	19,36	19,27	19,17	19,11	19,05			
2009		19,84	19,75	19,68	19,61	19,56		
2010			19,84	19,77	19,67	19,59	19,53	
2011				20,12	20,02	19,96	19,93	19,87
2012					20,62	20,56	20,50	20,46
2013						20,73	20,74	20,69
2014							21,27	21,22
2015								21,94

D'après fichiers détail EAR -2006-2013

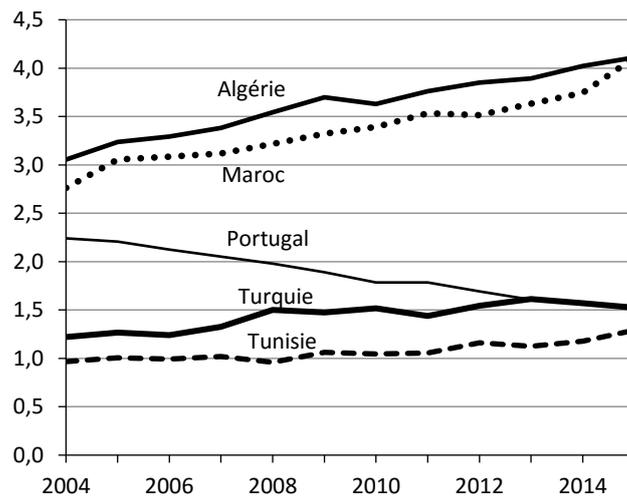
Les variations en ligne, en très légère diminution, sont pratiquement stables et les variations en colonne sont régulières, de l'ordre de 1,5 point en cinq ans. Il n'est donc pas étonnant que les deux courbes du graphique 3 soient quasiment confondues. On peut donc appliquer la même méthode que précédemment et estimer les proportions pour les quatre années entourant celles correspondant aux EAR (graphiques 3 et 4).

Graphique 3 – La proportion d'enfants d'origine étrangère selon le mode de calcul (2004-2015 pour les campagnes et 2006-2013 pour les EAR)



D'après fichiers détail EAR -2006-2013

Graphique 4 – proportion d'enfants pour cinq pays d'origines (2004-2015)



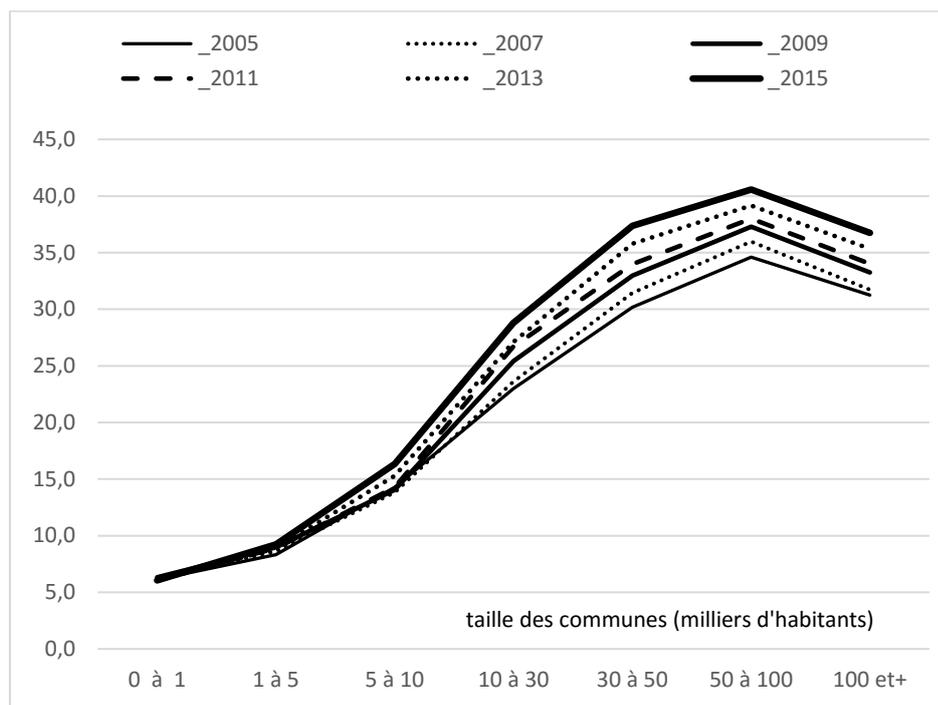
D'après fichiers détail EAR -2006-2013

L'influence de la taille des communes

Les quelque 36000 communes ont été réparties en 7 groupes selon leur taille. A chaque campagne correspond une courbe (graphique 5). La superposition des douze courbes traduit une grande régularité d'ensemble, ce qui plaide plutôt pour préférer les campagnes aux EAR.

Afin de prendre la mesure des écarts entre les deux modes de calcul, on donne (tableau 4) les valeurs fournies par les EAR et par les campagnes pour les 8 années communes aux deux traitements. Les écarts sont généralement infimes. Sur les 56 valeurs, seules 4 sont égales ou supérieures à un demi-point (la plus élevée est de 0,7 point). Comparés aux pourcentages, ces écarts sont donc négligeables, tout à fait indiscernables en représentation graphique.

Graphique 5 – proportion des enfants d'origine étrangère, superposition de 5 campagnes



D'après fichiers détail EAR 2006-2013

Tableau 4 – La proportion des enfants d'origine étrangère selon la taille des communes, selon les EAR (T 3a) et écarts avec les campagnes correspondantes (T 3b)

	2006	2007	2008	2009	2010	2011	2012	2013
Population	T3a - Valeurs des EAR							
<1	6,1	6,2	6,2	6,2	6,2	6,1	6,1	6,1
1 à < 5	8,5	8,6	8,7	8,7	8,9	8,9	9,0	9,1
5 à <10	13,7	13,8	14,0	14,3	14,6	14,9	15,2	15,4
10 à <30	23,3	24,0	24,6	25,4	26,0	26,4	26,9	27,5
30 à <50	30,9	31,8	32,4	32,9	33,6	34,2	34,9	35,7
50 à <100	35,0	35,8	36,4	36,9	37,4	37,9	38,4	39,1
>100	31,3	32,0	32,5	33,1	33,7	34,1	34,7	35,3
	T3b - Ecart entre campagnes et EAR							
<1	0,0	0,0	0,0	0,1	0,0	0,0	-0,1	-0,1
1 à < 5	0,1	0,0	0,0	0,2	-0,3	0,0	0,2	-0,2
5 à <10	-0,6	0,0	0,2	-0,3	0,5	-0,6	0,2	-0,1
10 à <30	-0,1	-0,4	0,2	0,0	0,2	0,3	-0,2	-0,4
30 à <50	0,3	-0,4	0,7	0,1	-0,4	-0,2	0,3	0,1
50 à <100	-0,2	0,2	0,2	0,4	-0,4	0,1	-0,3	0,1
>100	-0,2	-0,2	0,2	0,2	0,0	-0,1	0,0	0,0

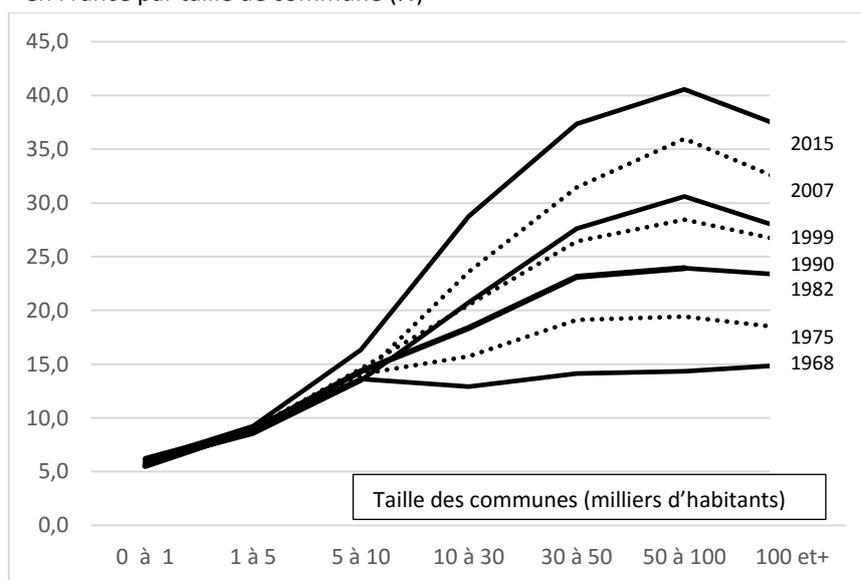
D'après fichiers détail EAR -2006-2013 Population en milliers d'habitants

De l'intérêt des séries très longues par raccordement aux anciens recensements

La proportion de jeunes d'origine étrangère est un exemple de ce que l'on peut faire lorsqu'on raccroche les anciens recensements (1962 à 1999) aux nouveaux (depuis 2006)⁵.

Le graphique 6 présente l'évolution de l'indicateur selon la taille des communes (même classement que précédemment) pour une période de près d'un demi-siècle (1968-2015). Chaque courbe correspond à une année de recensement (jusqu'en 1999) ou à une campagne. Dans les communes de moins de 10000 habitants, la proportion de jeunes d'origine étrangère a peu évolué au fil du temps. Au-delà, le même indicateur n'a cessé d'augmenter quasi-régulièrement avec toutefois un net ralentissement entre 1990 et 1999.

Graphique 6.- Évolution de la proportion de jeunes (0-17 ans) 1968-2015) d'origine étrangère en France par taille de commune (%)



D'après fichiers détail EAR -2006-2013

De toute évidence l'allure du graphique n'est pas banale. Il confirme s'il le fallait l'intérêt de disposer de séries longues. Il s'agit là d'un graphique concernant l'ensemble de la population immigrée, une population qui s'est profondément transformée au cours du temps. En 1968, l'indicateur était indépendant de la taille des communes, du moins au-dessus de 10000 habitants. Peu à peu, avec l'arrivée d'immigrés venus notamment du Maghreb et de l'Afrique subsaharienne, l'indicateur s'est accru rapidement dans les villes, plus spécialement dans les communes de banlieue de taille moyenne, d'où un bosselage des courbes de plus en plus prononcé.

Bien évidemment, il est possible de suivre les mêmes séries pour chaque grand groupe d'immigrants.

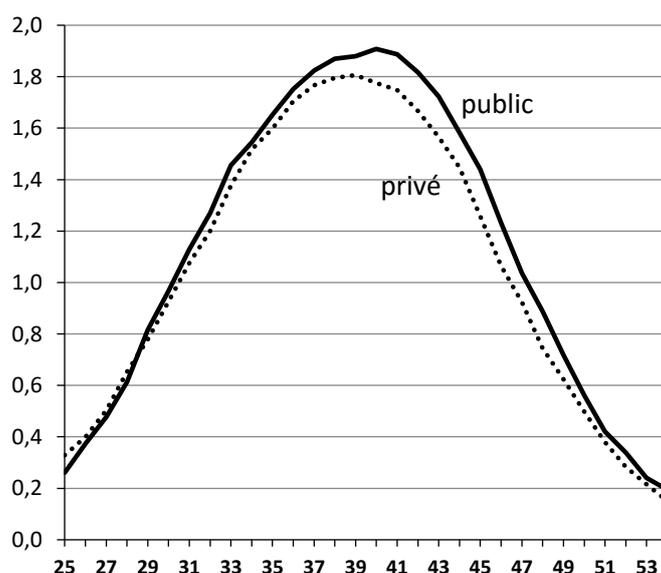
⁵Bernard Aubry et Michèle Tribalat – Les jeunes d'origine étrangère - Commentaire n°126, été 2009

C - La taille de la famille, selon le statut de la femme, secteur public ou privé (2004-2015)

On s'intéresse à la taille des famille et l'on cherche à savoir si les femmes travaillant dans le secteur public ont un nombre d'enfants à charge différent, en niveau et en évolution, de celles qui travaillent dans le secteur privé en CDI. L'analyse se fait par âge détaillé.

Le graphique 7 montre que toutes choses égales par ailleurs, les femmes du secteur public ont dans leur foyer un peu plus d'enfants que leurs collègues du privé, l'écart étant particulièrement significatif autour de 40 ans. En revanche, avant 30 ans la courbe s'inverse et au-delà de 50 ans, les écarts entre les deux catégories de femmes s'estompent, quoique, en valeur relative, il n'y ait pas semble-t-il de grand changement.

Graphique 7 : nombre d'enfants de moins de 18 ans selon l'âge de la mère en 2015

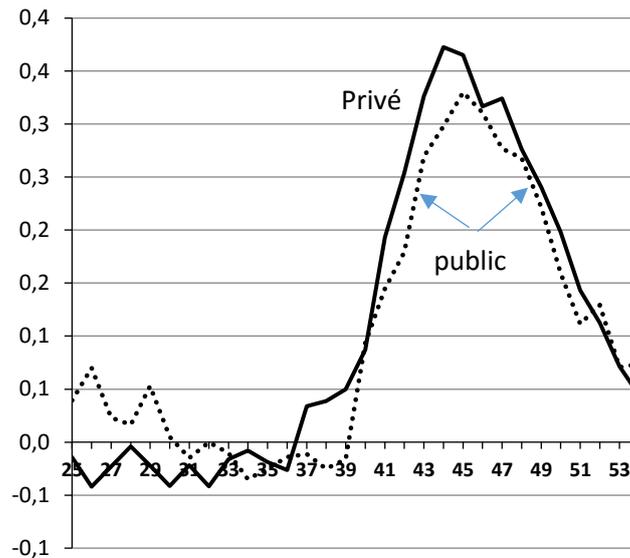


D'après fichiers détail EAR -2006-2013

Mais, en l'occurrence, c'est la variation dans le temps qui nous intéresse surtout et l'on constate (graphique 8) à la fois un clivage fort selon l'âge, mais aussi selon le statut. Quel que soit le statut, avant 40 ans, les variations de 2004 à 2015 sont faibles tandis qu'au-delà, et notamment autour de 45 ans, on assiste à une « explosion » du nombre d'enfants par femme.

Grosso modo, avant 35 ans le nombre d'enfants augmente plus vite chez les femmes du secteur public, mais chez les plus âgées, c'est l'inverse qui se produit.

Graphique 8- Variation de l'indicateur –2004-2015 _ secteurs public et privé



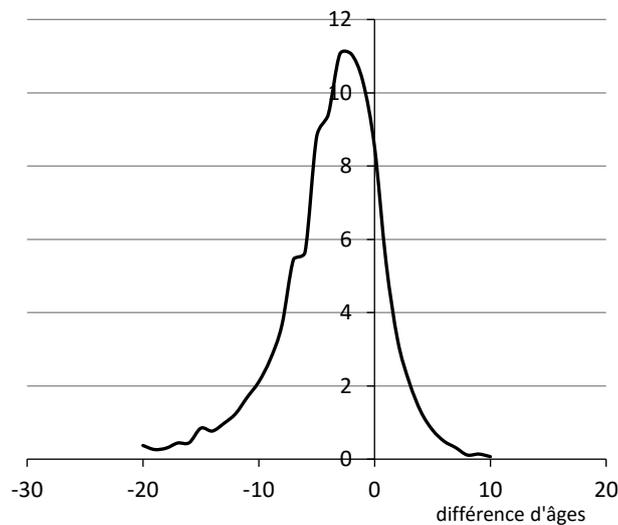
D'après fichiers détail EAR -2006-2013

La régularité des courbes, notamment celles du graphique 7 et du graphique 8 après 40 ans, plaide en faveur de l'utilisation des campagnes (les mêmes graphiques réalisés à partir des EAR limiteraient la période d'étude à 8 ans contre 12 pour les campagnes).

D - L'écart d'âge entre conjoints : les hommes nés en 1935 (2004-2015)

On regarde maintenant l'écart d'âge entre les membres des couples. Par convention on se fixe un âge (ou plutôt une année de naissance, à savoir 1935). Il s'agit donc d'hommes ayant environ 75 ans au milieu de la période d'étude. On suit le même cheminement que précédemment et on présente un graphique donnant la distribution des écarts en 2015.

Graphique 9 : les écarts d'âge en 2015



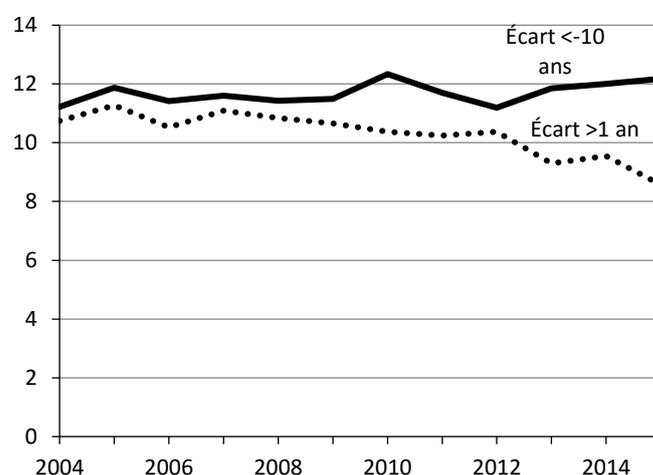
D'après fichiers détail EAR -2006-2013

L'écart modal est de 4 ans, mais la symétrie par rapport à l'âge modal n'est pas parfaite du fait de l'inégalité des sexes tant en ce qui concerne les comportements que la mortalité.

Pour le suivi temporel, on choisit deux groupes ayant au départ à peu près la même importance, à savoir d'une part les hommes ayant un conjoint 10 ans plus jeune et d'autre part ceux ayant un conjoint plus âgé de plus d'un an.

Les deux courbes restent assez proches l'une de l'autre jusqu'en 2012 (exception toutefois pour le point 2010), puis elles s'éloignent franchement l'une de l'autre avec, en 2015, un écart de plus de trois points. En l'occurrence le propos n'est pas de chercher à expliquer la différence, il est de montrer comment un phénomène qui n'aurait sans doute pas été perçu par l'exploitation des EAR, l'est par un traitement statistique des campagnes qui, chacune, forme un échantillon représentatif de la population du pays.

Graphique 10 – Chez les hommes nés en 1935, évolution de la proportion de conjoints ayant moins de dix ans - plus de 1 an (2004-2015)



D'après fichiers détail EAR -2006-2013

Conclusion

Pour aller plus loin, il faudrait examiner de façon systématique les domaines de validité de la « règle » qui voudrait que l'analyse par campagne soit supérieure à celle par EAR. Cela impliquerait de fixer des critères empiriques permettant la décision.

Les résultats favorables obtenus, notamment les écarts entre les deux modes de calcul en 2008 et 2015 pour l'ensemble de la métropole, le seraient-ils s'ils avaient été calculés par région, puisqu'a priori les

fichiers régionaux sont représentatifs ? Qu'en serait-il pour des territoires mixtes, c'est-à-dire des ensembles géographiques de grande taille, mais dont la représentativité ne serait pas assurée. Il n'est pas interdit d'aller voir.

Quoi qu'il en soit, une piste intéressante et facile à mettre en œuvre pour estimer de façon empirique le caractère significatif ou non des évolutions sur la période de 12 ans consisterait à comparer les évolutions sur l'ensemble de la période à la distribution des fluctuations annuelles.