

CONSEIL D'ORIENTATION DES RETRAITES

Séance plénière du 22 octobre 2013 à 14h30

« Les outils statistiques et les systèmes d'information sur les retraites »

Document N°9

<i>Document de travail, n'engage pas le Conseil</i>

**Un nouveau modèle pour estimer annuellement
des indicateurs « tous régimes de retraite confondus » : ANCETRE**

Patrick Aubert et Bruno Ducoudré, pour la DREES

CNAV, Retraite et Société n° 64, décembre 2012

f a i t s e t chiffres

Un nouveau modèle pour estimer annuellement des indicateurs « tous régimes de retraite confondus » : ANCETRE

Patrick Aubert, secrétariat général du COR
Bruno Ducoudré, OFCE, département analyses et prévisions

À NOTER

Les deux auteurs travaillaient au sein du bureau « Retraites » de la Direction de la recherche, des études, de l'évaluation et des statistiques (Drees) au moment de la rédaction de cet article.

Le système de retraite français se compose de plus d'une trentaine de régimes de base ou complémentaires différents, et la plupart des retraités cumulent des pensions de plusieurs régimes. Or, le besoin d'information porte souvent sur l'ensemble « tous régimes confondus ». Pour y répondre, la Drees produit plusieurs types de données statistiques : l'échantillon interrégimes de retraités (EIR), qui est une base individuelle « tous régimes » ; l'enquête annuelle auprès des caisses de retraite (EACR) qui regroupe des données agrégées portant sur les principaux régimes. Aucune de ces deux sources n'est cependant totalement satisfaisante à elle seule : les données de l'EIR sont très riches mais disponibles tous les 4 ans seulement ; l'EACR est réalisée chaque année pour collecter les données portant sur l'année précédente, mais ne contient aucune information sur la dimension « interrégimes ».

L'estimation annuelle de grandeurs « tous régimes » ne peut donc se faire que par une synthèse de sources statistiques diverses. Jusqu'à récemment cette estimation ne concernait que quelques indicateurs (nombre de retraités et pension moyenne de droit direct).

Elle était réalisée avec une méthode relativement fruste : le nombre de retraités « tous régimes » était calculé en divisant le nombre total de pensions versées (obtenu avec l'EACR) par le nombre moyen de pensions par retraité, lui-même calculé à partir du dernier EIR disponible et prolongé de façon *ad hoc*. Cette méthode comportait plusieurs inconvénients : les extrapolations *ad hoc* n'étaient pas toujours satisfaisantes, et toute l'information disponible n'était pas utilisée (par exemple la démographie). Chaque nouvel indicateur nécessitait aussi de développer une nouvelle méthode.

Pour toutes ces raisons, la Drees a récemment développé une nouvelle approche pour produire des indicateurs « tous régimes », fondée sur un modèle de microsimulation statique : le modèle ANCETRE (actualisation annuelle par calage pour l'estimation tous régimes des effectifs de retraités). Le modèle porte sur l'ensemble des régimes de base et complémentaires du système de retraite français. Il intègre l'ensemble des droits – droits propres et droits dérivés – versés par ces régimes. Il a été développé à partir de techniques de clonage et de calage, en couplant les données individuelles de l'EIR avec les données agrégées de l'EACR. De nouveaux indicateurs tous régimes ont ainsi été révisés ou développés avec ANCETRE (Drees, 2011), par exemple, les effectifs de retraités de droit direct et les effectifs de primo-liquidants d'un droit direct depuis 2004. Les estimations tous régimes par sexe et classe d'âge sont désormais cohérentes avec les estimations tous régimes globales d'une part, et les effectifs et montants moyens par caisses de retraite d'autre part.

Cette étude présente le principe général de fonctionnement du modèle ANCETRE. Elle illustre l'utilisation originale d'un modèle de microsimulation dans le cadre d'une production statistique courante d'indicateurs annuels et s'organise en deux parties. On présente tout d'abord la mécanique globale d'ANCETRE, puis on expose les principaux résultats et tests empiriques de validation du modèle.

Clonage et calage

Le principe du modèle est de « combiner » au mieux les sources statistiques, en conservant à la fois la richesse de l'EIR, notamment pour ce qui concerne la structure de pension, et le caractère actualisé de l'EACR.

La méthode consiste à conserver dans ANCETRE la structure de base des données individuelles du dernier EIR disponible et à construire un « pseudo-EIR » représentatif d'une année de référence plus récente¹. Plus précisément, les données individuelles d'ANCETRE sont issues de la réplication des données de l'EIR. « Réplication » n'implique pas pour autant reproduction à l'identique : on souhaite en effet en modifier la structure afin d'adapter les concepts et le champ de la base, et prendre en compte les évolutions survenues depuis l'année de base de l'EIR.

Dans un premier temps, certaines informations individuelles sont modifiées pour rendre les données plus représentatives de la nouvelle année de référence. Plus précisément, il s'agit de « décaler » la structure par âge de l'EIR, pour l'appliquer à des générations plus jeunes. Ainsi, on suppose par exemple que les individus de la génération 1944 (qui ont

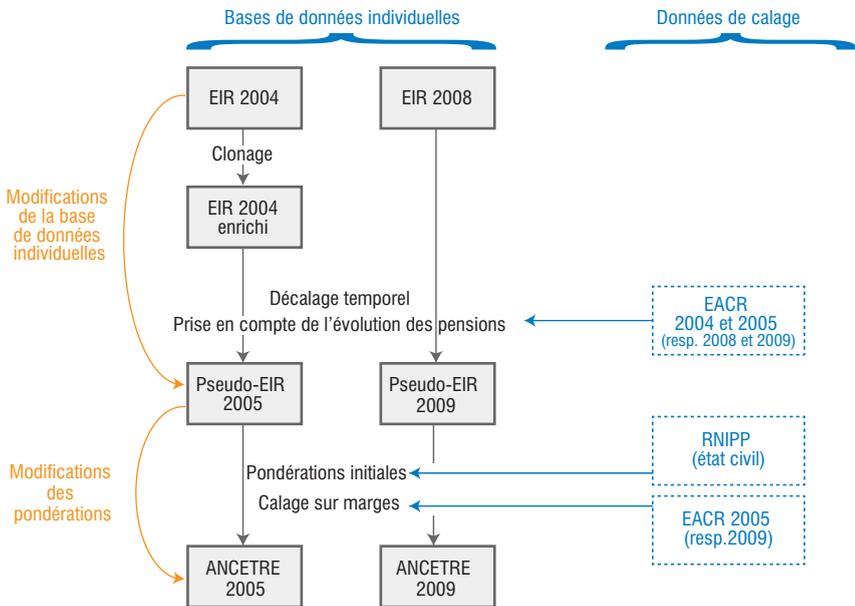
1. Par exemple, on transforme l'EIR 2004 en « pseudo-EIR 2005 ». Bac, *et al.* (2008) utilisent une méthode similaire pour l'actualisation du modèle de micro-simulation Myriade.

60 ans) dans l'EIR 2004 sont représentatifs des individus de la génération 1945, qui ont 60 ans en 2005. La méthode implique également de « cloner » certains individus pour représenter des individus absents de la base de données initiales, soit parce que leur génération ne faisait pas partie de l'échantillon de l'EIR de référence, soit pour prendre en compte la montée en charge de certains dispositifs nouveaux.

Dans un deuxième temps, on modifie la structure de pondération de l'échantillon afin de le rendre représentatif des informations démographiques les plus récentes, puis le « pseudo-EIR » est calé² sur les données actualisées (effectifs et montants de pension) des régimes de retraites, issues de l'EACR. On obtient en *output* des bases de données annuelles composées de pseudo-individus représentatifs de l'ensemble des retraités.

Le schéma 1 résume la mécanique d'ensemble de construction des bases ANCETRE, pour les exemples d'ANCETRE 2005 (construit à partir de l'EIR 2004) et d'ANCETRE 2009, (construit à partir de l'EIR 2008). La suite de la partie 1 détaille cette mécanique.

Schéma 1



Concepts et champ des retraités

Les « pseudo-EIR » reposent sur les mêmes concepts et le même champ que l'EIR : les données individuelles en sont issues. Ces données sont cependant articulées avec celles de l'EACR, et la cohérence de deux sources doit être assurée. Malgré leur proximité – en termes de concepts et de champ – le rapprochement des deux sources nécessite de définir au plus juste le champ d'étude retenu, certains écarts pouvant subsister du fait des différences de modes de collecte.

2. Par une procédure de calage sur marges.

Les concepts et le champ ne sont pas non plus rigoureusement identiques d'une vague de l'EIR à l'autre : en l'occurrence, ils ont légèrement évolué entre l'EIR de 2004 et celui de 2008. Afin d'éviter des ruptures de série entre les pseudo-EIR construits à partir de l'EIR de 2004 et ceux construits à partir de l'EIR de 2008, il est nécessaire d'homogénéiser les concepts et le champ entre les deux vagues de cet échantillon. En pratique, ce sont ceux de l'EIR de 2008 qui sont retenus.

Quelques définitions de base

La définition des retraités inclut l'ensemble des retraités des régimes présents dans l'EIR (encadré 1) au 31 décembre de l'année *N*.

Les liquidations sous la forme d'un versement forfaitaire unique ne sont pas considérées comme des pensions de retraite.

ENCADRÉ 1

L'échantillon interrégimes de retraités (EIR)

L'EIR est un panel qui permet notamment d'observer l'évolution du nombre et de la composition des retraités, et de reconstituer le montant moyen de pension par retraité tous régimes pour un échantillon anonyme d'individus, en rapprochant les données des différents régimes français légalement obligatoires. Ces informations sont collectées tous les quatre ans depuis 1988.

L'EIR de 2008 portant sur les pensions au 31 décembre 2008 est la sixième vague de cette opération statistique. Il comprend tous les individus de l'échantillon qui perçoivent une pension de droit direct ou de droit dérivé d'une caisse de retraite, qu'ils vivent en France ou à l'étranger. La quasi-totalité des organismes de retraite obligatoire (régimes de base, régimes spéciaux et régimes complémentaires légalement obligatoires) sont interrogés, soit 74 régimes au total. Les retraites issues de régimes supplémentaires non obligatoires sont en revanche exclues du champ de l'opération, ainsi que les revenus provenant de l'épargne individuelle volontaire et les pensions versées par des systèmes de retraite étrangers.

La sélection des individus se fait sur la base de la date de naissance. L'EIR de 2004 portait sur les générations âgées de 54 ans ou plus (soit 130 00 retraités). Le champ de l'EIR 2008 a été étendu aux générations plus récentes (252 00 individus au total), pour mieux prendre en compte les possibilités de départs anticipés dans certains régimes spéciaux.

Pour chacun des individus de l'EIR, les caisses de retraite renseignent :

- les caractéristiques individuelles du retraité (sexe...)
- la nature et le montant des prestations (régime de retraite, montant des avantages perçus...)
- les conditions de liquidation des droits à la retraite (âge de liquidation...)
- le détail des éléments de calcul du montant de pension (salaire de référence, durées validées...).

La liquidation d'un droit est appréciée à la date d'effet (entrée en jouissance) de ce droit. Si cette date est postérieure au 31 décembre de l'année N , ou si le retraité est décédé avant le 31 décembre de l'année N , la pension n'est pas comptabilisée au titre de l'année N .

Pensions de retraite et pensions d'invalidité

Dans les régimes de la fonction publique et certains régimes spéciaux, les pensions d'invalidité et les pensions de retraite sont traitées dans le même système de gestion. Par conséquent, il est très difficile d'établir une distinction nette entre le moment où une pension, liquidée à l'origine pour un motif d'invalidité, correspond à une prestation d'invalidité et le moment où elle correspond à une prestation de vieillesse. À l'inverse, les pensions d'invalidité sont gérées dans un système distinct du système de retraite pour les affiliés des régimes général et alignés. Elles prennent fin à l'âge de 60 ans et ne donnent lieu qu'alors à la liquidation d'une pension de retraite.

Par souci de comparaison, on a donc considéré comme des pensions de retraite toutes les pensions d'invalidité des anciens fonctionnaires dont l'âge est supérieur à l'âge « normal » de départ à la retraite. Lorsque l'âge est inférieur, les pensions d'invalidité sont en revanche exclues du champ de l'EIR 2008. Le champ de l'EIR 2008 est légèrement plus large que celui de l'EACR : le premier inclut certains titulaires de pensions d'invalidité de la fonction publique et des régimes spéciaux âgés de 50 à 59 ans (lorsque leur âge « normal » de départ à la retraite est inférieur à 60 ans), alors que le second ne retient que les anciens invalides de 60 ans et plus. Le concept retenu dans ANCETRE pour l'intégration des pensions d'invalidité est rigoureusement identique à celui de l'EIR 2008.

Clonage et décalage temporel

Le « clonage » est une étape spécifique à la constitution des pseudo-EIR construits à partir de l'EIR 2004. Cette vague de l'EIR est en effet constituée de moins de générations que l'EIR 2008 (tableau 1). Afin d'éviter les ruptures de séries lors du passage à l'EIR 2008, on a « cloné » des individus de l'EIR 2004, afin de créer des groupes d'individus représentatifs de générations initialement absentes de l'EIR 2004. La base de données individuelles d'ANCETRE pour l'année 2004 est donc constituée des pensions des individus présents dans l'EIR 2004, et des pensions de « pseudo-individus » construits à partir de l'EIR 2004, afin de reproduire une structure par âge proche de celle de l'EIR 2008.

Tableau 1 > Principales différences de plan de sondage entre les EIR 2004 et 2008

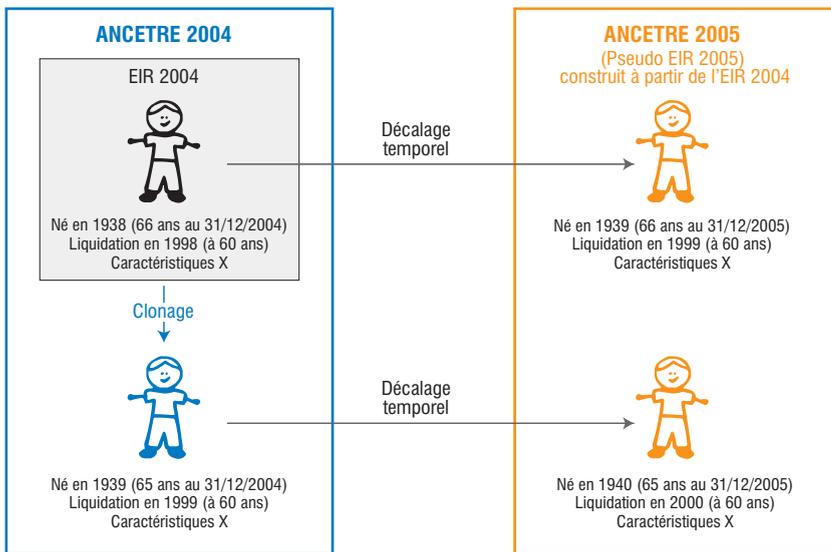
Âge à la date d'observation	EIR 2004	EIR 2008
Moins de 50 ans	Aucune génération	Une génération sur deux
54 à 66 ans	Une génération sur deux	Toutes les générations

La création d'un pseudo-EIR consiste ensuite à répliquer la structure par âge de l'EIR, pour l'appliquer à des générations plus jeunes, en décalant uniformément toutes les variables temporelles de l'EIR (années de naissance, de liquidation, etc.). Comme ce « décalage temporel » correspond à celui de l'année d'observation de référence, il a pour effet de maintenir constant l'âge des individus dans l'échantillon. L'hypothèse sous-jacente est que les caractéristiques des individus en termes de carrières, d'acquisition des droits à la retraite et de comportements de départ restent stables pour des générations proches.

Le clonage : création des générations absentes dans l'EIR 2004

Pour refléter l'ensemble des générations âgées de 55 à 65 ans, comme dans l'EIR 2008, il s'agit de recréer les générations nées les années impaires à partir des générations nées les années paires. Par exemple, pour créer la génération née en 1939, on duplique toutes les observations des individus de la génération 1938, en conservant uniquement les pensions liquidées au plus tard le 31 décembre 2003, soit à 65 ans. On augmente ensuite arbitrairement d'une unité, pour les individus clonés, l'année de naissance, l'année de cessation d'activité et l'année de liquidation de leur pension. Les autres caractéristiques sont conservées à l'identique (schéma 2). La méthode de clonage abaisse le taux d'individus retraités retrouvés dans une génération impaire, puisqu'il n'est pas tenu compte de la mortalité³.

Schéma 2



Un clonage spécifique : les départs anticipés pour carrière longue

Le dispositif de départ anticipé pour carrière longue a été introduit par la réforme de 2003. Les individus peuvent liquider leur pension avant 60 ans dès lors qu'ils réunissent certaines conditions portant sur la durée d'assurance validée tous régimes, la durée d'assurance cotisée et l'âge de début de carrière. Du fait de son entrée en vigueur en 2004, on ne dispose dans l'EIR 2004 que du premier flux de liquidants cette année-là. Dans ANCETRE 2004, on crée donc des individus partant à la retraite anticipée pour les générations d'année de naissance impaire (1945 et 1947) à partir des individus ayant liquidé un droit direct en 2004 et ayant bénéficié du dispositif de retraite anticipée pour les générations 1944 et 1946. Ainsi, les départs l'année des 59 ans sont clonés à partir des

3. Cette sous-estimation est néanmoins corrigée, au moins en partie, par le calage sur les effectifs agrégés par génération fournis par l'EACR (cf. *infra*).

départs l'année des 60 ans (intervenues entre janvier et septembre⁴) et des départs l'année des 58 ans (intervenues entre octobre et décembre). Les départs l'année des 57 ans sont clonés à partir des départs l'année des 58 ans⁵.

Lorsque les pensions sont dupliquées, seule la génération est modifiée. L'ensemble des pensions (base et complémentaires) de chaque individu liquidé en 2004 est conservé. Le clonage est donc légèrement différent de celui appliqué pour les âges de liquidation supérieurs à 60 ans. Notons que les données de l'EIR ne permettent pas de discriminer les individus qui auraient potentiellement rempli les conditions de départ dans le cadre du dispositif avant 2004, si celui-ci avait existé avant cette date, des autres (le nombre de trimestres validés à 16 ou 17 ans par exemple). On peut donc s'attendre à cloner un peu trop d'individus en 2004⁶.

Pour les années suivantes, les départs anticipés sont aussi ventilés selon qu'ils ont lieu l'année donnée ou les années précédentes, l'EIR 2004 n'ayant pas d'historique sur cette mesure. On utilise les données de l'EIR 2008 pour attribuer aléatoirement une année de liquidation aux individus. Ce point est nécessaire, car le calage est ensuite effectué sur le stock de retraités, mais aussi sur le flux de liquidants.

Les spécificités du décalage temporel pour l'année 2009

Certaines évolutions réglementaires survenues en 2009 font que la création d'un pseudo-EIR pour 2009 ne peut pas se faire par simple décalage temporel de la totalité des observations de l'EIR 2008.

Les conditions d'accès au dispositif de départ anticipé pour carrière longue sont en effet durcies en 2009. Les pensions des pseudo-individus liquidant en 2009 en départ anticipé et ne remplissant pas ces nouvelles conditions sont exclues d'ANCETRE. On tient aussi compte du nouveau seuil de liquidation sous la forme d'un versement forfaitaire unique à l'Ircantec. Enfin, on intègre la condition d'âge rétablie à 55 ans pour obtenir une pension de réversion au régime général et dans les régimes alignés.

Évolution des pensions

En général, la pension moyenne des retraités augmente d'année en année. Cette augmentation ne reflète pas l'évolution de l'avantage principal perçu, individuellement, par chaque retraité. Elle provient en partie des revalorisations annuelles appliquées par chaque caisse de retraite, mais s'explique aussi par un effet de composition lié au renouvellement de la population des retraités. Celui-ci se traduit par l'arrivée de nouveaux retraités ayant généralement eu des carrières plus favorables et disposant en moyenne

4. L'EIR est constitué d'individus exclusivement nés en octobre. En 2004, date de création du dispositif de départ anticipé pour carrière longue, les individus nés en 1944 n'ont donc pu bénéficier du dispositif qu'entre janvier et septembre de la même année.

5. En d'autres termes, pour chaque retraité de la génération 1946 (qui a 58 ans en 2004) parti en 2004 dans le cadre du dispositif de départ anticipé pour carrière longue, on crée un clone né en 1947. La date de liquidation est conservée à l'identique entre le clone et son modèle, ce qui revient à considérer un âge à la liquidation d'un an plus jeune pour le clone.

6. Cet exercice a bien sûr ses limites : en particulier, on ne vérifie pas si, pour tous les individus clonés, les informations individuelles portant sur les durées validées sont bien cohérentes avec les conditions d'éligibilité. Par ailleurs on ne dispose pas d'information dans l'EIR sur l'âge de début de carrière, et on ne peut donc pas savoir si la condition d'éligibilité correspondante est bien vérifiée pour les générations clonées.

de pensions plus élevées, et par le décès de retraités plus âgés percevant des pensions plus faibles en moyenne.

Pour rendre compte de ce mécanisme, dans ANCETRE, l'évolution de la pension moyenne intervient par différents canaux. Premièrement, pour les régimes les plus importants, on dispose de données par génération, sexe et année sur le montant moyen de l'avantage principal de droit direct pour l'ensemble des retraités des régimes et pour les nouveaux bénéficiaires. Pour une année $N > 2004$, les pensions individuelles $pens_{a,N}$ des retraités d'âge a sont revalorisées suivant l'évolution des pensions moyennes $\overline{pens}_{a,s,N}$ par âge et par sexe s pour ces régimes :

$$pens_{a,s,2006} = pens_{a,s,2004} \times \frac{\overline{pens}_{a,s,2006}}{\overline{pens}_{a,s,2004}}$$

Pour les nouveaux retraités de l'année, on revalorise les pensions de la même façon en appliquant les évolutions moyennes par âge d'une année sur l'autre. En 2005, la pension d'un nouveau bénéficiaire $pens_{60,s,2005}^{L(2005)}$ âgé de 60 ans et ayant liquidé cette année-là est égale à la pension observée en 2004 corrigée de l'évolution moyenne entre 2004 et 2005 de la pension moyenne des liquidants du même âge dans la caisse :

$$pens_{a,s,N}^{L(N)} = pens_{a,s,N-1}^{L(N-1)} \times \frac{\overline{pens}_{a,s,N}^{L(N)}}{\overline{pens}_{a,s,N-1}^{L(N-1)}}$$

Pour les années postérieures à 2005, les évolutions sont chaînées. Ainsi, pour 2006, les individus représentatifs des retraités âgés de 61 ans et dont la pension a été liquidée l'année des 60 ans (en 2005) ont leur pension revalorisée par rapport à 2004 comme celle des liquidants âgés de 60 ans en 2005 par rapport aux liquidants âgés de 60 ans en 2004, puis selon l'évolution de la pension moyenne entre 2005 et 2006 pour l'ensemble des retraités âgés de 61 ans dans le régime :

$$pens_{61,s,2006}^{L(2006)} = pens_{61,s,2004}^{L(2003)} \times \frac{\overline{pens}_{60,s,2005}^{L(2005)}}{\overline{pens}_{60,s,2004}^{L(2004)}} \times \frac{\overline{pens}_{61,s,2006}}{\overline{pens}_{61,s,2005}}$$

Il s'agit de capter au mieux les évolutions de pension à âge donné provenant des évolutions de carrière entre générations.

Pour les autres (petits) régimes, les pensions sont revalorisées en fonction des revalorisations légales appliquées par chaque régime. On s'attend donc à sous-estimer au fil des ans la pension moyenne. Mais, pour tous les régimes participant à l'EACR, les pensions moyennes totales⁷ sont calées sur les pensions moyennes issues de l'EACR (cf. *infra*). Cette étape de repondération par calage sur marges permet donc de corriger la sous-estimation.

Calage sur marges

Les pondérations initiales sont calculées pour tenir compte de la démographie (annexe), puis le calage sur marges permet de caler les agrégats (nombre de retraités et pensions moyennes) principalement sur les données de cadrage fournies par les caisses de retraite dans le cadre de l'EACR. Le calage est effectué avec la macro Calmar.

7. Plus précisément, on cale l'avantage principal de droit direct, sauf pour l'Agirc où on cale la pension moyenne totale.

Les données

Les données de l'EACR utilisées incluent les effectifs de retraités, les effectifs de nouveaux bénéficiaires ainsi que les montants moyens de pension, ventilés par sexe et par génération.

Pour les régimes absents de l'EACR, on a utilisé les données publiées dans le rapport annuel de la Commission des comptes de la Sécurité sociale portant sur le nombre de pensions de droit direct et dérivé.

ENCADRÉ 2

L'enquête annuelle auprès des caisses de retraite (EACR)

L'EACR, dont les résultats sont disponibles depuis 2004, recueille des informations sur les effectifs de retraités et les pensions versées par la plupart des caisses de retraite*. Les caisses participant à cette enquête sont :

- régimes de base : Cnav, MSA salariés et non-salariés, RSI ;
- régimes spéciaux : service des retraites de l'État, CNRACL, SNCF, RATP et Cniieg ;
- régimes complémentaires : Arrco, Agirc, Ircantec ;
- régimes complémentaires obligatoires des artisans et commerçants et des exploitants agricoles.

Les données recueillies portent sur les retraités de droit direct ou de droit dérivé (retraités bénéficiant d'une pension de réversion) vivants au 31 décembre de l'année. L'enquête collecte notamment des informations sur les montants moyens de pensions, le nombre de retraités ayant perçu ces pensions, le nombre de liquidants, ainsi que les motifs de liquidation (inaptitude, etc.). L'enquête recueille également des informations ventilées par âge et par sexe pour les principales variables.

Des demandes complémentaires à l'enquête ont été menées récemment afin de collecter des « séries historiques » d'effectifs de bénéficiaires d'une pension de droit direct et des montants moyens de pension correspondant ventilés par âge et, pour certaines caisses, par lieu de naissance (nés en France/nés à l'étranger). La collecte de données sur les pensions de réversion a aussi été progressivement améliorée : une collecte de tableaux ventilés par âge a été ajoutée à partir de 2008 pour la Cnav et le SRE, et étendue progressivement aux autres caisses en 2009. Les ventilations par âge ont aussi été étendues, pour toutes les variables, de 20 ans à 95 ans à partir de 2009, contre 50 à 85 ans auparavant.

* Des travaux sont en cours pour élargir l'enquête à de nouvelles caisses de retraite.

Le calage

Les « séries historiques » de l'EACR ont permis un calage par génération des droits directs depuis 2004 pour les principales caisses de retraite de base et pour les caisses de la fonction publiques (annexes 2 et 3 de Aubert et Ducoudré, 2011, pour une présentation complète des marges retenues). Le calage par génération pour les caisses de retraites complémentaires et les régimes spéciaux a été ajouté à partir de 2008.

Les marges contenant un faible effectif (inférieur à 2 000 retraités) n'ont pas été conservées, afin de faciliter le calage. On a utilisé la méthode linéaire tronquée (Sautory, 1993). Les bornes pour le ratio de poids (rapport du poids après calage sur le poids avant calage) ont été choisies par itérations successives. Pour les années 2004 à 2007, les bornes sont plus lâches que pour les années suivantes, le clonage ne permettant pas de rendre compte parfaitement de la montée des départs anticipés pour carrières longues (tableau 2).

Tableau 2 > Caractéristiques du calage

Année	Nombre de marges	Ratio de poids minimum	Ratio de poids maximum
2004	1 003	0,1	5,79
2005	1 078	0,1	11,36
2006	1 091	0,1	12
2007	1 085	0,1	14
2008	1 748	0,2	6,8
2009	2 122	0,2	7,1

Concernant la réversion, ANCETRE ne permet pas actuellement de calculer de façon satisfaisante des effectifs de bénéficiaires d'un droit dérivé tous régimes confondus avant 2008, pour deux raisons : premièrement, les séries historiques ne permettent pas de remonter jusqu'en 2004 pour l'Arcco ; deuxièmement, une partie des pensions de réversion versées par l'Arcco a été imputée dans l'EIR 2004, sur la base de données de cadrage issues de l'ancien système d'information de l'Arcco, ce qui crée une rupture avec les données collectées dans l'EIR 2008.

Quelques résultats et tests de validation empirique

Cette partie présente quelques résultats permettant d'apprécier la « qualité » du modèle. On confronte pour cela divers estimateurs issus d'ANCETRE à d'autres données de cadrage.

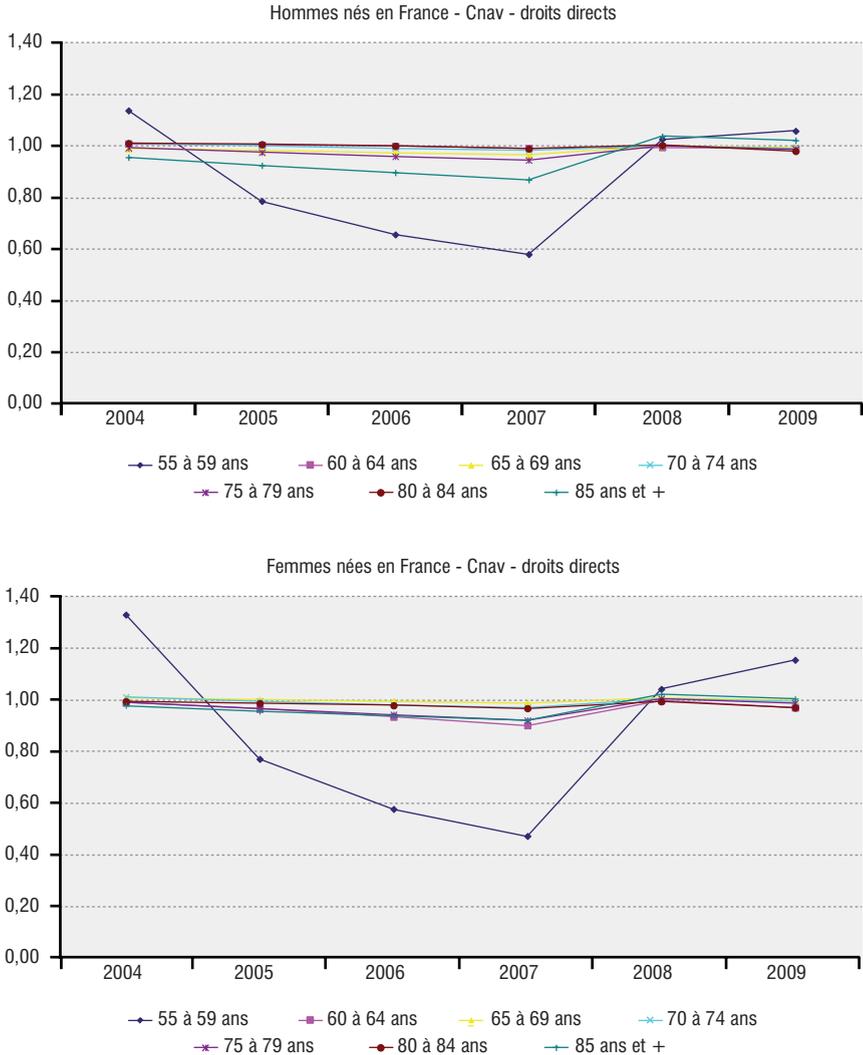
Les effectifs de retraités : avant et après calage

Si l'on compare, pour les retraités de la Cnav titulaires d'un droit direct, les effectifs pondérés avant calage issus d'ANCETRE avec les effectifs issus de l'EACR, on obtient, pour les nés en France hommes et femmes, des résultats très satisfaisants dans toutes les classes d'âge entre 60 et 85 ans (graphique 1).

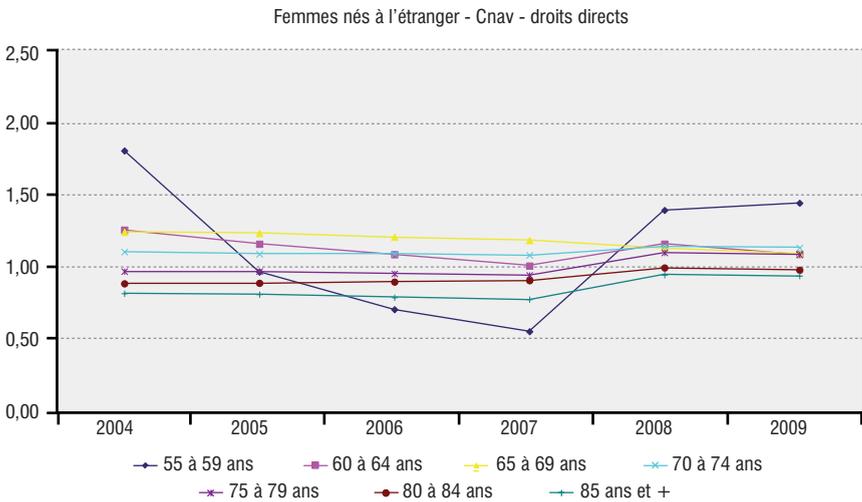
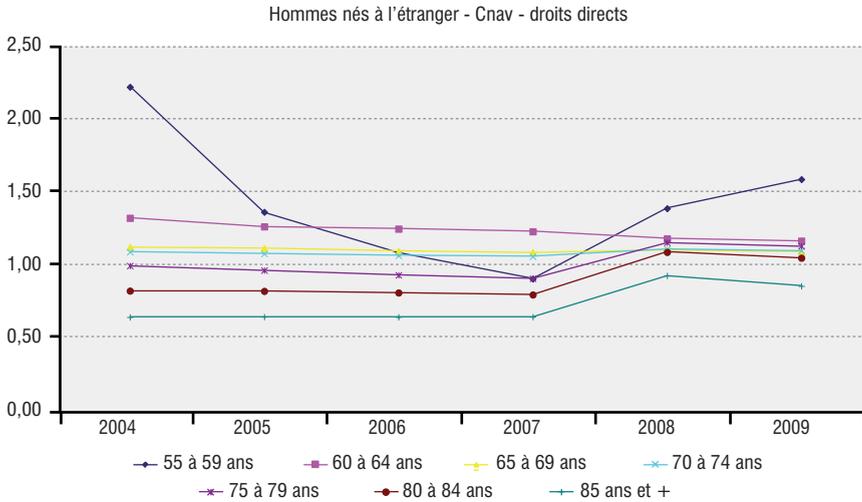
En revanche, les effectifs d'hommes retraités âgés de moins de 60 ans sont surestimés de 14 % en 2004, ce qui peut s'expliquer par la difficulté à bien répliquer les conditions d'éligibilité au dispositif de départ anticipé pour carrière longue. La montée en charge de ce dispositif implique une sous-estimation des départs anticipés les années suivantes, pour les hommes (-42 % en 2007) comme pour les femmes (-52 % en 2007) : le clonage seul ne permet pas de rendre compte de façon satisfaisante de la montée en charge du dispositif. Par contre, le passage à l'EIR 2008 permet une meilleure estimation initiale de ces effectifs, la surestimation en 2009 pouvant provenir du durcissement des conditions d'accès au dispositif, qui ne serait qu'imparfaitement prise en compte dans la constitution du pseudo-EIR pour cette année 2009.

Concernant les nés à l'étranger, l'estimation initiale des effectifs est moins précise. Avant 2008, les effectifs de retraités de plus de 80 ans sont sous-estimés, tandis que ceux de 60 à 74 ans sont surestimés. Les départs anticipés sont aussi fortement surestimés, les nés à l'étranger ayant généralement des durées validées plus courtes⁸.

Graphique 1 > Ratio entre les effectifs de retraités de droit direct calculés avec ANCETRE (pondérations initiales) et les effectifs tirés de l'EACR pour la Cnav



8. On aurait pu envisager d'appliquer des conditions restrictives de durée validée pour le clonage avant 60 ans, en excluant par exemple les individus à faible durée validée tous régimes. Cependant, cette information n'est pas toujours disponible dans l'EIR 2004. Par ailleurs, la surestimation est moins importante les années suivantes, les effectifs étant sous-estimés pour les femmes nées à l'étranger à partir de 2006.



Les départs à la retraite des femmes âgées de 60 à 64 ans en 2007 sont par ailleurs légèrement sous-estimés, ce qui peut provenir de la hausse progressive du taux d'activité des femmes, qui induit une augmentation du pourcentage de femmes ayant un droit à pension de retraite au fil des générations.

Ancienne méthode de calcul des indicateurs « tous régimes » versus ANCETRE

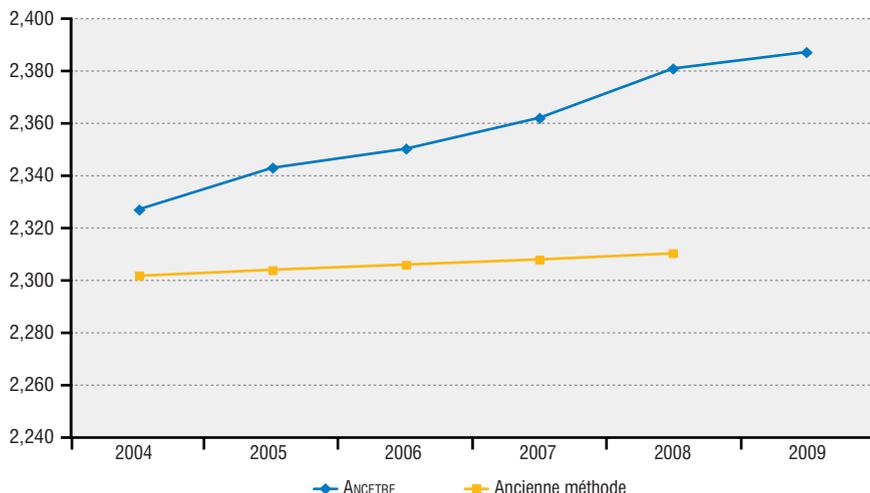
L'ancienne méthode de calcul des effectifs de retraités tous régimes confondus consistait à partir de la somme des effectifs de chaque régime. Celle-ci mesure le nombre total des pensions servies et non le nombre global des retraités couverts par ces régimes. Pour éliminer les doubles comptes, le nombre des pensions versées était divisé par une estimation, pour chaque année d'observation, du nombre moyen de pensions versées par retraité.

Par exemple, on estimait qu'un retraité percevait en moyenne 2,3 pensions de droit direct en 2007. Or, les caisses de retraite servent 33,5 millions de pensions de droit direct cette même année. L'effectif total de retraités pour 2007 était donc estimé à 14,5 millions (33,5 divisé par 2,3).

La pension mensuelle moyenne tous régimes confondus était ensuite obtenue en rapportant la somme des versements effectués mensuellement par l'ensemble des caisses de retraite au nombre global de retraités calculé précédemment.

Ces calculs étaient effectués par la Drees en utilisant l'EIR et l'EACR. La somme des pensions versées tous régimes confondus était calculée d'après les données de l'EACR, alors que le nombre moyen de pensions versées par retraité était extrait de l'EIR de 2004.

Graphique 2 > Comparaison entre le nombre moyen de pensions de droit direct par retraité calculé avec l'ancienne méthode et celui estimé dans ANCETRE



Sources : Drees, ANCETRE, EACR, EIR 2004.

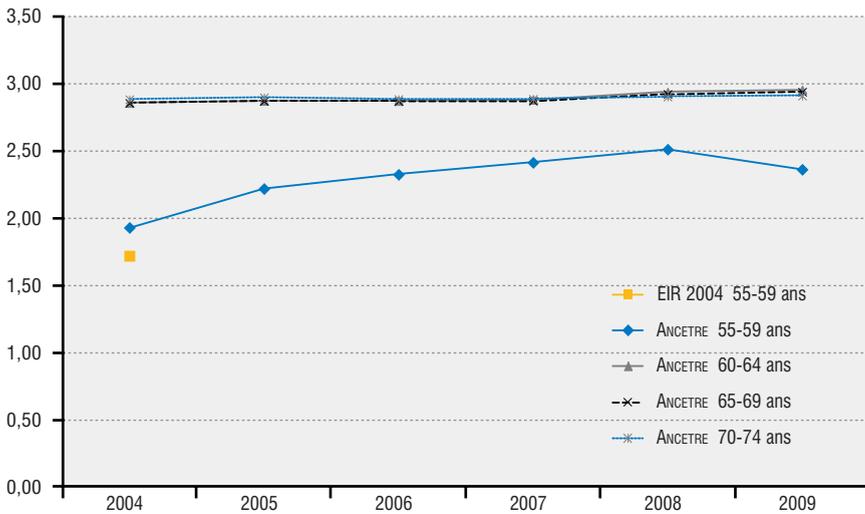
Cette méthode reposait fortement sur l'estimation du nombre moyen de pensions par retraité. Pour les années postérieures à 2004, une tendance *ad hoc* était appliquée pour rendre compte de l'augmentation de ce nombre. Si l'on compare cette estimation avec celle obtenue par ANCETRE, on constate que la tendance⁹ était largement sous-estimée entre 2004 et 2008 (graphique 2). Cela s'explique par la montée en charge des départs anticipés pour carrière longue : les retraités liquidant dans le cadre de ce dispositif liquident souvent plusieurs pensions¹⁰ (graphique 3).

9. Le niveau initial estimé en 2004 est également différent : l'écart correspond à la différence de concept et de champ (traitement des pensions d'invalidité des anciens fonctionnaires et prise en compte des retraités de moins de 55 ans, cf. *supra*).

10. Les bénéficiaires de ce dispositif sont en grande majorité d'anciens salariés du privé, et liquident donc à ce titre au moins deux ou trois pensions de droit direct (base et complémentaire). À l'inverse, avant 2004, les retraités de moins de 60 ans étaient principalement des retraités de régimes spéciaux intégrés, et ne percevaient donc souvent qu'une seule pension.

La méthode de construction d'ANCETRE, par le biais des étapes de clonage et de calage, permet de mieux rendre compte de ce phénomène. Ainsi, on n'observe aucune rupture de série entre les nombres moyens de pensions par retraité pour les années 2007 (base ANCETRE construite à partir des données de l'EIR 2004) et 2008 (la base ANCETRE correspond alors exactement à l'EIR 2008, et fournit donc la « vraie » estimation pour cette année 2008).

Graphique 3 > Nombre moyen de pensions de droit direct par retraité selon la classe d'âge



Sources : Drees, ANCETRE, EACR, EIR 2004.

Basage sur l'EIR 2004 versus basage sur l'EIR 2008

Afin d'étudier la robustesse de notre méthodologie, on compare maintenant les résultats de l'EIR pour l'année 2008 avec ceux obtenus en créant, grâce au modèle ANCETRE, un pseudo-EIR 2008 à partir de l'EIR 2004. Cela permet d'analyser dans quelle mesure les indicateurs se déforment au fil du temps avec le calage.

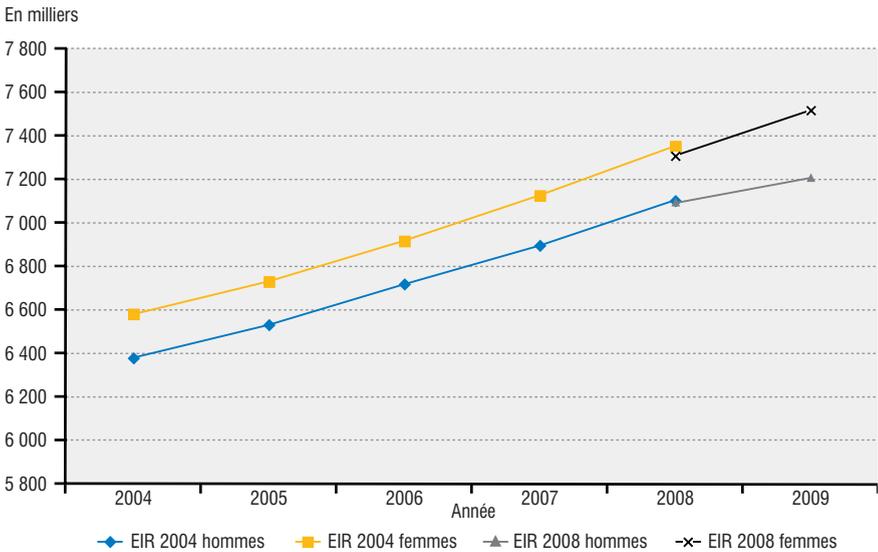
Si l'on compare les effectifs de retraités de droit direct tous régimes confondus, les effectifs calculés à partir d'ANCETRE construit avec l'EIR 2004 sont supérieurs de 0,4 % par rapport à ceux calculés avec l'EIR 2008 (graphique 4). Ce sont principalement les effectifs de femmes qui sont surestimés (+0,6 % contre +0,2 % pour les hommes). Les effectifs sont globalement surestimés pour les retraités les plus jeunes, et sous-estimés pour les retraités plus âgés : ils sont surestimés entre 55 et 69 ans et entre 75 et 79 ans pour les hommes, et entre 55 et 84 ans pour les femmes.

Ces écarts peuvent provenir de plusieurs facteurs, qui ne se cumulent pas nécessairement. Ainsi, la structure de polypension (à âge donné) pourrait se déformer au fil des générations. Dans ce cas, la prise en compte des évolutions démographiques, par

le biais des pondérations initiales calculées par génération à partir des données de l'état civil, ne suffit pas. D'autre part, le calage de l'EIR 2008 est plus précis, plus de caisses étant calées par générations. Ces deux facteurs sont difficilement discernables puisque le calage a un impact sur la structure de polypension. Cependant, les nombres moyens de pension par retraité restent globalement très proches si on les compare par sexe et classe d'âge (graphique 5). Les pensions moyennes par âge sont aussi globalement proches entre les deux méthodes : l'avantage principal de droit direct moyen est sous-estimé de seulement 0,1 % pour les hommes et de 0,4 % pour les femmes en 2008 (graphique 6).

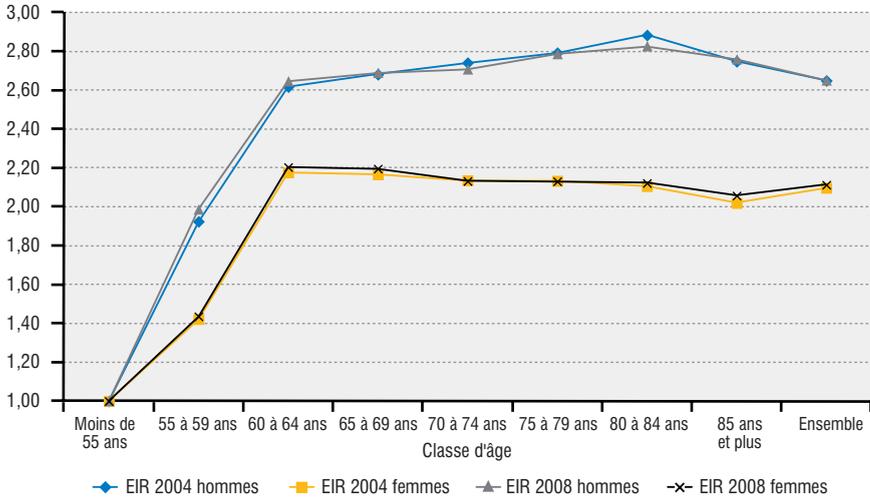
Finalement, la création d'un pseudo-EIR avec le modèle ANCETRE n'entraîne pas de dérive trop importante des indicateurs estimés par rapport à l'estimation plus précise obtenue avec l'EIR 2008. En particulier, l'écart observé en 2008 entre les deux méthodes pour ces indicateurs reste de faible ampleur, par comparaison à leurs variations annuelles (de 2007 à 2008 et de 2008 à 2009). Par ailleurs, un calage sur marges plus précis en 2008, puis pour les années 2009 et au-delà devrait permettre de réduire la dérive des indicateurs tous régimes en attendant le prochain EIR 2012.

Graphique 4 > Effectifs de retraités de droit direct tous régimes confondus, par sexe, selon la vague de l'EIR retenue dans le modèle ANCETRE



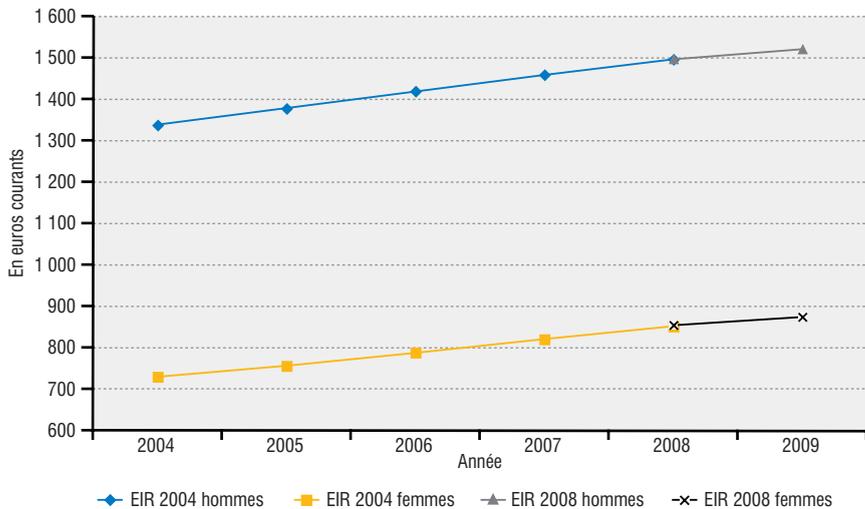
Note : la série « EIR 2004 » correspond aux pseudo-EIR annuels successifs construits à partir de cette vague de l'EIR, et la série « EIR 2008 » aux pseudo-EIR construits à partir de l'EIR de 2008.
 Source : Drees, ANCETRE.

Graphique 5 > Nombre moyen de pensions de droit direct par retraité selon la classe d'âge en 2008



Note : la courbe « EIR 2004 » correspond au pseudo-EIR 2008 construit à partir de l'EIR de 2004.
Source : Drees, ANCETRE.

Graphique 6 > Pension mensuelle moyenne de droit direct tous régimes confondus, par sexe, selon la vague de l'EIR retenue dans le modèle ANCETRE



Note : la série « EIR 2004 » correspond aux pseudo-EIR annuels successifs construits à partir de cette vague de l'EIR, et la série « EIR 2008 » aux pseudo-EIR construits à partir de l'EIR de 2008.
Source : Drees, ANCETRE.

Conclusion : limites et pistes d'amélioration du modèle

Une limite des résultats présentés ici est leur faible profondeur temporelle, qui réduit les possibilités d'analyse des évolutions au cours du temps. Le modèle ANCETRE peut cependant être adapté afin de construire des résultats tous régimes avant 2004. Aubert et Ducoudré (2011, annexe 5) présentent notamment une rétopolation entre 2000 et 2004.

L'élargissement du champ de l'EACR aux caisses ne participant pas actuellement à l'enquête permettra aussi une amélioration substantielle d'ANCETRE : homogénéisation des concepts statistiques, calage plus précis des effectifs par génération ou encore meilleure prise en compte de l'évolution des pensions moyennes au fil des générations. ANCETRE reste adapté essentiellement pour fournir des indicateurs qui sont issus des variables pour lesquelles il y a un calage sur les données de l'EACR, à savoir les données d'effectifs et de montants de pension. Bien sûr, rien n'empêche, techniquement, d'utiliser les bases ANCETRE pour calculer d'autres indicateurs tous régimes (proportions associées à certains dispositifs tels que la décote, la surcote, etc.). Mais ces indicateurs seraient alors plus fragiles, du fait de l'absence de calage spécifique sur les variables correspondantes. Pour élargir le champ des indicateurs tous régimes, il est toutefois envisageable d'ajouter de nouvelles marges de calage dans ANCETRE.

Dans le même ordre d'idée, ANCETRE n'est pas adapté pour prendre en compte correctement l'impact d'évolutions réglementaires, telles que la hausse de la durée d'assurance requise sur les comportements de départ à la retraite ou sur le niveau des pensions. En effet, lors du clonage des individus, les éléments constitutifs du calcul des droits à pension sont laissés à l'identique. Sous l'effet des réformes de ce type, il peut donc apparaître progressivement une dérive entre les résultats d'ANCETRE et les « vrais » résultats tous régimes.

La prise en compte des modifications structurelles liées à de telles réformes constitue l'une des pistes d'amélioration du modèle à l'avenir. Plusieurs pistes pour cela peuvent en effet être envisagées, même si elles demandent un lourd travail de modélisation, ce qui explique qu'elles n'aient pas été explorées dans la version actuelle du modèle. À titre d'exemple, la méthode d'actualisation des montants de pensions – dans la première étape de construction du modèle – pourrait être revue. Plutôt que de corriger directement le montant de pension individuel, on pourrait modifier d'abord, pour chaque individu, chacun des éléments constitutifs de ce montant (salaire de référence, durée validée, durée cotisée, etc.) en modélisant une « dérive » au fil des générations. Puis, on recalculerait le montant de pension en appliquant la législation propre à chaque année, au moyen d'un outil de simulation des montants de pension tous régimes, tel que l'outil Caliper de la Drees (Lerméchin, *et al.*, 2011).

Bibliographie

Aubert P., Ducoudré B., 2011, « Le modèle ANCETRE : actualisation annuelle par calage pour l'estimation tous régimes des effectifs de retraités », Drees, documents de travail, série *Sources et méthodes*, n° 24, septembre.

Bac C., Legendre F., Lorgnet J.-P., Marc C., Thibault F., 2008, « Recalage de l'enquête Revenus fiscaux et clonage des ménages : les deux prérequis à l'utilisation du modèle de micro-simulation Myriade », *Courrier des statistiques*, n° 123, janvier-avril.

Drees, 2011, « Les retraités et les retraites en 2009 », *Études et statistiques*, juin.

Lermechin H., Duc C., Burricand C., 2011, « Présentation et applications de l'outil Caliper (calcul interrégimes des pensions de retraite) », Drees, documents de travail, série *Études et recherche*, n° 111, novembre.

Sautory O., 1993, « La macro Calmar : redressement d'un échantillon par calage sur marges », Insee, documents de travail de la direction des Statistiques démographiques et sociales, n° F9310, novembre.



Annexe

Les pondérations initiales

Dans ANCETRE, les pondérations initiales sont calculées pour tenir compte de la démographie (informations issues des fichiers d'état civil de l'Insee, lesquels constituent également la base de tirage pour l'échantillon de l'EIR).

L'EIR est échantillonné dans le répertoire national d'identification des personnes physiques (RNIPP). Le tirage est effectué par sexe et par lieu de naissance (inscription dans la section « métropole » ou dans la section « hors métropole » du RNIPP) pour plusieurs générations. La sélection est en réalité réalisée selon le jour de naissance des individus, mais, si on fait l'hypothèse que le jour de naissance est aléatoire, cette sélection est analogue à un tirage avec probabilité d'inclusion uniforme au sein de chaque catégorie de sexe, lieu de naissance et génération.

La construction des pondérations initiales consiste, par analogie, à rendre les individus échantillonnés représentatifs de l'ensemble des individus de la même catégorie inscrits au RNIPP. Toutes les générations n'étant pas représentées dans l'EIR, des regroupements de générations R sont effectués (Aubert et Ducoudré, 2011, annexe 1). Le poids initial P d'un individu i de sexe s , appartenant à la génération g ¹¹ inscrit dans la section sec du RNIPP et présent dans l'échantillon ANCETRE s'écrit ainsi :

$$P_{i, g, s, sec} = \frac{\sum_{g \in R, s, sec} NB_{g, s, sec}}{\sum_{g, s, sec} nb_{g, s, sec}}$$

avec NB = nombre total d'individus de sexe s , appartenant à la génération g inscrit dans la section sec du *mipp* vivants au 31/12/N, et nb = nombre total d'individus dans ANCETRE de sexe s , appartenant à la génération g inscrit dans la section sec du RNIPP vivants au 31/12/N.

11. Il s'agit de la génération à l'issue de l'étape de décalage temporel.