

UN NOUVEL OUTIL STATISTIQUE POUR ESTIMER ANNUELLEMENT DES INDICATEURS « TOUS RÉGIMES DE RETRAITE CONFONDUS » : LE MODÈLE ANCETRE

Patrick AUBERT (), Bruno DUCOUDRÉ(**)*

() Insee, Département des études économiques d'ensemble*

*(**) OFCE*

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.

Introduction

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 –notamment en ce qui concerne la pension moyenne ou les effectifs de retraités– porte souvent sur l'ensemble « tous régimes confondus ». Pour répondre à ce besoin, 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, et est donc disponible plus rapidement, mais elle ne contient aucune information sur la dimension « inter-régimes ».

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, effectuée par la Drees, ne concernait que quelques indicateurs (nombre de retraités et pension moyenne de droit direct). Elle était de plus réalisée avec une méthode relativement fruste : on calculait un nombre de pensions moyen par retraité à partir du dernier EIR disponible, puis on imputait de façon ad-hoc une tendance d'évolution de ce ratio. Le nombre de retraités « tous régimes » était ensuite calculé en divisant le nombre total de pensions versées (obtenu avec l'EACR) par le nombre moyen de pensions par retraité calculé auparavant. Cette méthode comportait plusieurs inconvénients : les extrapolations ad-hoc ne sont pas toujours satisfaisantes, et toute l'information disponible n'était pas utilisée (par exemple, la démographie ou la montée en charge de certains dispositifs). Par ailleurs, il fallait développer une nouvelle méthode pour chaque nouvel indicateur que l'on souhaitait produire.

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). De nouveaux indicateurs tous régimes ont ainsi été révisés ou développés avec ANCETRE (Drees, 2011) : les effectifs de retraités de droit direct, les effectifs de primo-liquidants d'un droit direct et les montants moyens de pension tous régimes confondus depuis 2004, ou encore les effectifs bénéficiaires d'un droit dérivé tous régimes confondus depuis 2008. Des estimations tous régimes par sexe, classe d'âge et pour certaines générations sont désormais possibles et cohérentes avec les estimations tous régimes globales d'une part, et les effectifs et montants moyens par caisses de retraite d'autre part.

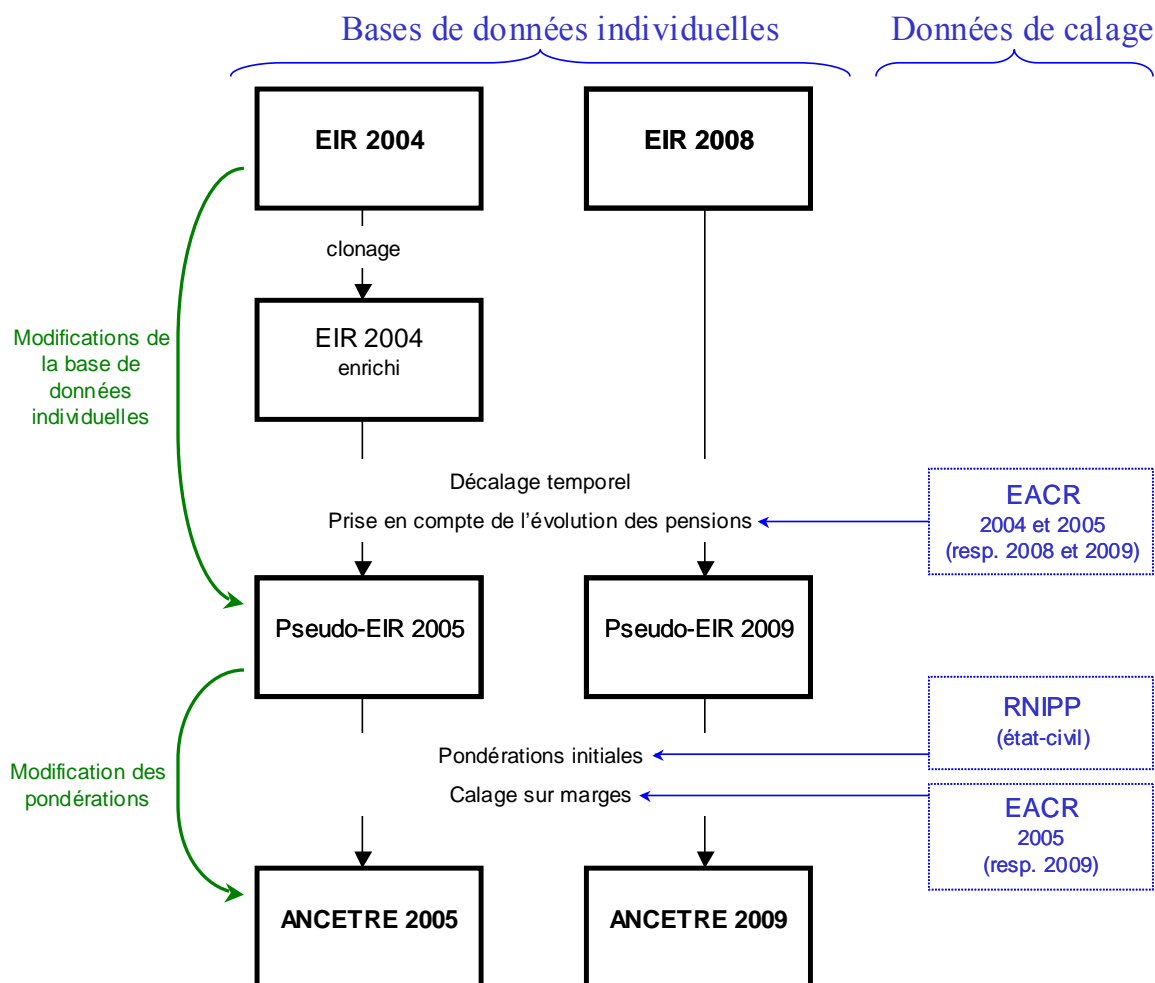
Cette étude présente le principe général de fonctionnement du modèle ANCETRE. Elle illustre ainsi l'utilisation originale d'un modèle de microsimulation dans le cadre d'une production statistique courante d'indicateurs annuels.

Pour donner dès à présent une idée de la méthode employée, le principe est de « combiner » au mieux les sources statistiques, en conservant à la fois la richesse de l'EIR et le caractère actualisé de l'EACR. Plus précisément, il s'agit de conserver dans ANCETRE la structure de base des données individuelles du dernier EIR disponible et de construire un « pseudo-EIR » représentatif d'une année de référence plus récente¹. Cette construction s'opère en deux étapes. 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 répliquer la structure par âge de l'EIR, pour l'appliquer à des générations plus jeunes. 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. Ainsi, on suppose par exemple que les individus de la génération 1944 (qui ont 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 qu'ils n'étaient pas échantillonnés dans l'EIR de référence (par exemple, le champ de l'EIR 2004 n'inclut qu'une génération sur deux et ne couvre pas la totalité du champ des retraités), soit pour prendre en compte la montée en charge de certains dispositifs nouveaux (notamment les retraites anticipées pour carrière longue à partir de 2004). Dans un deuxième temps, on modifie la structure de pondération de l'échantillon afin de le rendre représentatif des données agrégées les plus récentes. On calcule pour cela des pondérations initiales individuelles qui tiennent compte des informations démographiques connues et actualisées, puis le « pseudo-EIR » est calé² sur les données actualisées des régimes de retraites, issues de l'EACR. Ces données retenues pour le calage portent sur les effectifs de retraités, les effectifs de nouveaux bénéficiaires ainsi que sur les montants moyens de pension, ventilés par sexe et par génération. On obtient en output des bases de données annuelles composées de pseudo-individus représentatifs de l'ensemble des retraités. Ces bases permettent des calculs tous régimes actualisés annuellement en termes d'effectifs et de montants, ainsi que des ventilations cohérentes avec les estimations tous régimes globales (par exemple des estimations selon le lieu de résidence des retraités).

Le schéma ci-dessous 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 l'étude détaille cette mécanique. Elle est organisée en quatre parties. On présente tout d'abord les modifications des données individuelles : méthode de clonage, de décalage temporel et prise en compte des revalorisations de montants d'une année sur l'autre. Puis on décrit la méthode de pondération : construction des pondérations initiales, données utilisées pour le calage et résultats du calage sur marge. La troisième partie est consacrée à la présentation d'une méthodologie complémentaire, permettant de construire des bases « réropolées » pour les années antérieures à 2004. Dans la quatrième partie, on présente enfin les principaux résultats et les principaux tests empiriques de validation du modèle.

¹ Par exemple, on transforme l'EIR 2004 en « pseudo-EIR 2005 ».

² Par une procédure de calage sur marges.



1. Modifications des données individuelles : la création des « pseudos-EIR »

Pour ce qui concerne la structure de polypension, le modèle ANCETRE se fonde sur l'information issue de l'EIR. Plus précisément, les bases de données individuelles construites par le modèle le sont par réplique des données de l'EIR. « Réplique » n'implique pas pour autant reproduction à l'identique. On souhaite en effet modifier légèrement la structure des données individuelles, afin d'adapter les concepts et le champ de la base, et afin de prendre en compte certaines modifications réglementaires survenues depuis l'année de référence de l'EIR.

Cette modification des données individuelles peut recouvrir un changement de certaines informations pour des individus de la base de données, mais aussi la création de nouveaux individus dans cette base.

1.1. Homogénéisation des concepts et du champ des retraités

Les « pseudo-EIR » créés par le modèle ANCETRE reposent, par construction, sur les mêmes concepts et le même champ que l'EIR, puisque les données individuelles en sont issues. Ces données sont cependant articulées avec celles de l'EACR, et la mise en cohérence de deux sources doit donc être assurée³. Malgré leur proximité –en termes de concepts et de champ– le

³ On rappelle que l'EIR est un échantillon quadriennal en panel qui regroupe des informations individuelles sur les pensions versées par les caisses de retraite (une part importante des retraités percevant plusieurs pensions versées par des régimes différents). L'EACR collecte des informations annuelles exhaustives sur l'ensemble des retraités des principaux régimes et sur les montants de pension versés.

rapprochement de deux sources d'information nécessite en effet de définir au plus juste le champ d'étude retenu, certains écarts pouvant subsister du fait des différences de modes de collecte.

Par ailleurs, 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 donc nécessaire, au préalable, 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.

1.1.1. Quelques définitions de base

La définition des retraités inclut l'ensemble des retraités des régimes présents dans l'EIR au 31 décembre de l'année N (champ géographique Métropole + Dom + Tom + étranger, pour le lieu de naissance et pour le lieu de résidence). Ne sont pas comptabilisées, pour les retraités nés ou résidants en France, les pensions versées par des systèmes de retraite étrangers, ainsi que par quelques régimes spéciaux de faible effectif (Port autonome de Strasbourg, Opéra de Paris, CNBF⁴, etc.) et par les régimes de retraite supplémentaire facultative (par exemple la retraite facultative des commerçants ex-« Organic complémentaire »). Les liquidations sous la forme d'un versement forfaitaire unique ne sont pas considérées comme des pensions de retraite.

La liquidation d'un droit est appréciée à la date d'effet (entrée en jouissance) de ce droit⁵. 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. Par contre, si le retraité décède entre le 1^{er} janvier de l'année N+1 et la date de la collecte des données pour l'année N, la pension est comptabilisée au titre de l'année N⁶.

1.1.2. Enrichissement de l'EIR 2004 à partir de l'EIR 2008

Il est fréquent d'observer un décalage entre la date d'entrée en jouissance d'une pension et la liquidation administrative de celle-ci. Un certain nombre de pensions ayant une entrée en jouissance en 2004 étaient ainsi absentes de l'EIR 2004, car connues des caisses de retraite après la collecte des données constitutives de l'EIR 2004. On utilise l'EIR 2008 pour récupérer ces pensions liquidées tardivement⁷. Les montants de pension, exprimés en euros en valeur au 31 décembre 2008 sont corrigés des revalorisations de pensions intervenues entre le 1^{er} janvier 2005 et le 31 décembre 2008.

Pour les présent/présent, on peut soit conserver les lignes de 2004, soit celles de 2008. Du fait de possibles révisions de la pension entre 2004 et 2008, on a fait le choix de conserver les lignes de 2004.

Concernant l'ARRCO et l'AGIRC, les retraités peuvent cumuler plusieurs pensions au sein même de ces régimes. Les lignes de l'EIR 2008 sont donc agrégées⁸ avec comme date de liquidation retenue celle de la caisse principale. Pour compléter l'EIR 2004, on récupère l'ensemble des lignes de 2008, avant agrégation et dont l'année d'entrée en jouissance n'excède pas 2004. L'agrégation par ligne est

⁴ Le CNBF a été intégré dans l'EIR 2008, mais pas dans l'EACR. Si les caisses présentes dans l'EIR ne sont pas interrogées dans le cadre de l'EACR, d'autres sources sont mobilisées lorsque c'est possible.

⁵ Les données administratives de certains régimes contiennent parfois d'autres dates en plus de la date d'entrée en jouissance, telles que la date de premier versement de la pension ou la date de traitement du dossier administratif. Or celles-ci peuvent différer de la date d'entrée en jouissance du droit.

⁶ En pratique, cette convention n'est pas toujours parfaitement vérifiée, les bases de données administratives des caisses ne permettant pas toujours les redressements statistiques nécessaires. Cela peut par exemple être le cas lorsque la liquidation d'une pension avec date d'entrée en jouissance en N intervient en N+1, et que le retraité décède entre la date de liquidation administrative et la date de collecte des données.

⁷ A cette fin, on restreint l'EIR 2004 à l'échantillon commun avec l'EIR 2008 pour éviter un biais de collecte. Les individus de la génération née en 1906, uniquement échantillonnés en 2004, sont néanmoins conservés pour rendre compte des pensions des plus âgés. Pour cette génération, on peut légitimement penser que les pensions liquidées tardivement sont négligeables.

⁸ Dans l'EIR 2008, les données sont regroupées par croisement individu X régime de retraite. Les diverses caisses ARRCO sont ainsi, par exemple, regroupées et agrégées au sein du « régime » ARRCO unique.

ensuite effectuée sur l'ensemble de l'EIR 2004 avec les mêmes critères de sélection des variables que ceux de l'EIR 2008 pour ces deux caisses.

1.1.3. 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. Une conséquence est qu'il est très difficile pour ces régimes 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.

Par souci de comparaison avec les régimes du privé, les bénéficiaires d'une pension d'invalidité dans un régime de la fonction publique ou dans un régime spécial sont, au dessus d'un certain âge, inclus dans le champ des retraités. En effet, les pensions d'invalidité prennent fin à l'âge de 60 ans et donnent lieu à la liquidation d'une pension de retraite dans les régimes – général et alignés –, alors qu'elles demeurent des pensions d'invalidité dans les régimes de la Fonction publique. Par convention, 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 des pensions de retraite, et donc exclues du champ de l'EIR⁹.

Cette convention est appliquée d'une manière légèrement différente selon la source des données. Dans l'EACR, les pensions d'invalidité des régimes de la Fonction publique sont intégrées dès lors que la personne a atteint au moins 60 ans, et exclues si elle est plus jeune. Dans l'EIR, l'information disponible au niveau individuel permet d'évaluer de manière un peu plus fine l'âge « normal » de départ à la retraite : ce dernier tient compte du statut de chacun (qui détermine l'âge d'ouverture des droits) et peut donc valoir, selon les cas, 50, 55 ou 60 ans pour les cas les plus fréquents¹⁰. Le champ de l'EIR est donc légèrement plus large que celui de l'EACR puisqu'il inclut, de surcroît, certains titulaires de pensions d'invalidité de la Fonction publique et des régimes spéciaux âgés de 50 à 59 ans. Le concept retenu dans ANCETRE pour l'intégration des pensions d'invalidité est rigoureusement identique à celui de l'EIR 2008.

Par ailleurs, les pensions de réversion issues d'une pension d'invalidité sont considérées comme des pensions de réversion de retraite quel que soit l'âge du bénéficiaire.

1.2. 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. Elle ne porte que sur les retraités âgés de 54 ans et plus en 2004, tandis que l'EIR 2008 inclut aussi les retraités de moins de 50 ans (une génération sur deux est sondée de la génération 1956 à la génération 1974). Par ailleurs, chaque génération est sondée entre 54 et 66 ans, contre une sur deux dans l'EIR 2004. 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. *In fine*, l'objectif est de reproduire une structure par âge proche de celle de l'EIR 2008.

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 d'autres termes, il s'agit de décaler uniformément toutes les variables temporelles de l'EIR (année de naissance, année de liquidation, etc.) Comme ce « décalage temporel » correspond à celui de l'année d'observation de référence, il a pour effet de maintenir

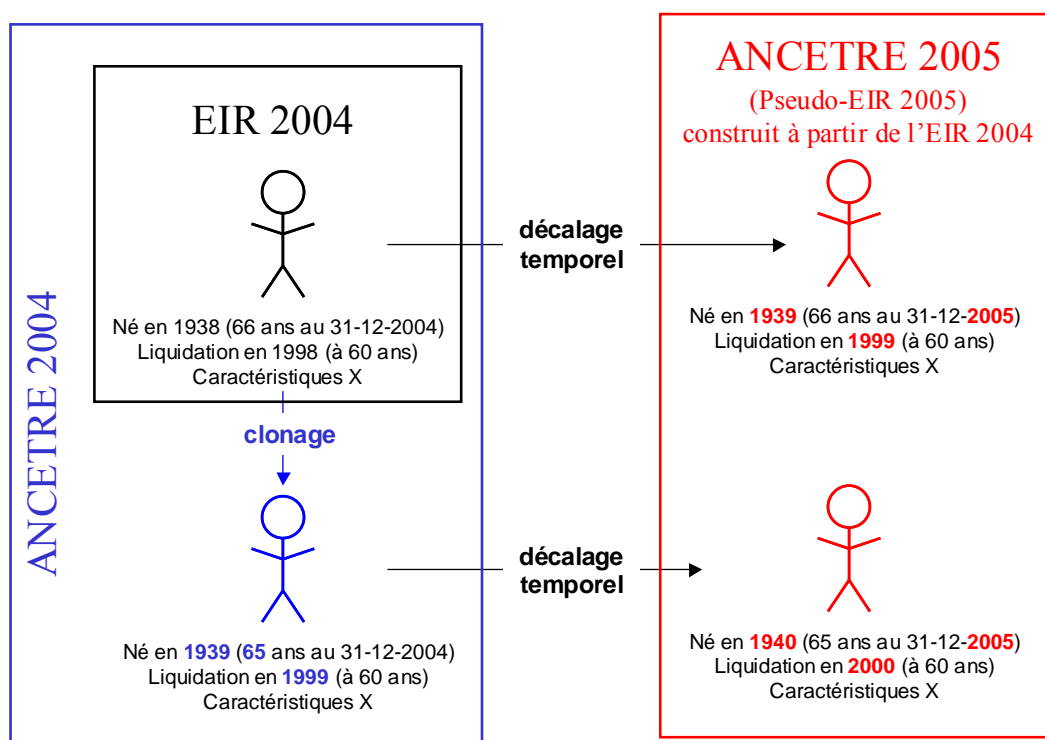
⁹ Dans l'EIR de 2004, toutes les pensions d'invalidité des régimes spéciaux étaient retenues dans la base de données, et donc considérées comme des pensions de retraite, quel que soit l'âge du bénéficiaire à la date d'observation.

¹⁰ Avant la réforme de 2010, certains corps de la Fonction publique ont un âge d'ouverture des droits compris entre 50 et 55 ans ou entre 55 et 60 ans selon les cas. Pour les générations parties à la retraite récemment, l'année d'ouverture des droits est souvent renseignée dans l'EIR 2008 et permet de calculer l'âge d'ouverture des droits.

constant l'âge des individus dans l'échantillon. Ainsi, on suppose par exemple que les individus de la génération 1944 (qui ont 60 ans) dans l'EIR 2004 sont représentatifs des individus de la génération 1945, qui ont 60 ans en 2005. 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. ANCETRE est donc constitué de bases de données individuelles annuelles entre 2004 et 2009.

1.2.1. 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 31/12/2003, 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'identique¹¹ (cf. schéma). La méthode de clonage abaisse le taux d'individus retraités retrouvés dans une génération impaire, puisqu'il n'est pas tenu compte de la mortalité¹².



Pour rendre compte des retraités de moins de 54 ans, qui sont principalement pensionnés de la Fonction Publique et des régimes spéciaux, on crée des générations paires de 32 à 52 ans à partir de la génération née en 1950. Par exemple, la génération ayant 52 ans est constituée des retraités de la génération 1950 de la Fonction Publique et des régimes spéciaux ayant une année d'entrée en jouissance strictement inférieure à 2003 dans l'EIR 2004.

¹¹ Les individus ayant à la fois une pension de droit direct et une pension de droit dérivé dans la même caisse, pensions dont l'année d'entrée en jouissance est différente, peuvent être clonés partiellement : si un pensionné a liquidé un droit dérivé en 2004 mais bénéficie d'une pension de droit direct dont l'année d'entrée en jouissance est inférieure, seule la partie portant sur la pension de droit direct est conservée pour l'individu cloné.

¹² Les générations utilisées pour le clonage sont néanmoins âgées de 66 ans au plus. Pour la génération née en 1938, les tables de mortalité de l'Ined indiquent une mortalité de 1,3% l'année des 66 ans pour les individus ayant survécu jusqu'à cet âge. Par ailleurs, le calage par génération sur les effectifs agrégés fournis par l'EACR permet de corriger en partie ce biais en rehaussant les effectifs pondérés. D'autre part, le calage est aussi effectué sur les montants moyens de pension : si le montant de pension individuel est corrélé à la mortalité (Aubert & Christel Andrieux, 2010), le calage permet de corriger en partie du biais lié à la mortalité différentielle.

1.2.2. Un clonage spécifique : les départs anticipés pour carrière longue

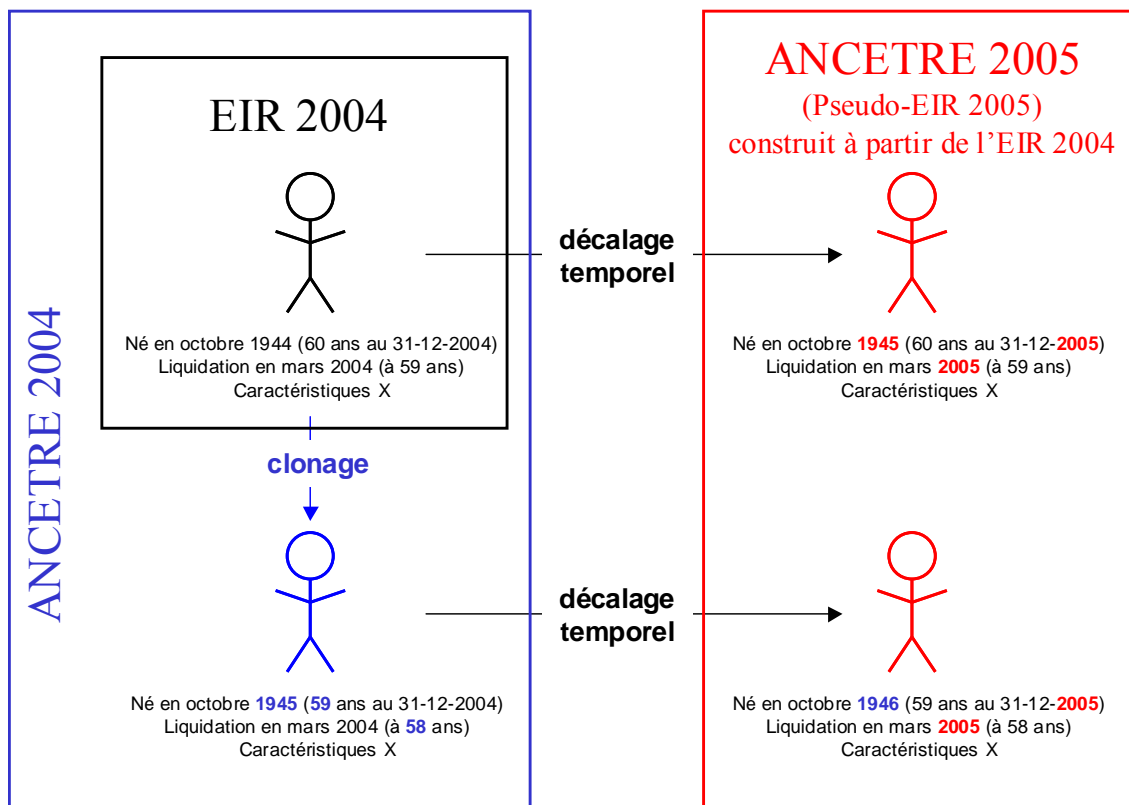
Le dispositif de départ anticipé pour carrière longue a été introduit par la réforme de 2003. Les individus peuvent liquider leur pension avant 60 ans dès lors qu'ils réunissent certaines conditions portant sur la durée d'assurance validée tous régimes, la durée d'assurance cotisée et l'âge de début 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 départs l'année des 60 ans (intervenues entre janvier et septembre¹³) 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 ans¹⁴.

Lorsque les pensions sont dupliquées, seule la génération est modifiée. L'ensemble des pensions de chaque individu (pensions de base et complémentaires) sont conservées si elles ont été liquidées en 2004. Le clonage est donc légèrement différent de celui appliqué pour les âges de liquidation supérieurs à 60 ans (cf. schéma), pour lesquelles le problème de création du dispositif en 2004 seulement ne se posait pas. 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 2004¹⁵.

¹³ 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.

¹⁴ 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.

¹⁵ 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.



Pour les années suivantes, il faut aussi ventiler les départs anticipés selon qu'ils ont lieu l'année donnée ou les années précédentes, l'EIR 2004 n'ayant pas d'historique sur cette mesure : tous les retraités de moins de 60 ans en 2004 ayant liquidé dans le cadre de ce dispositif l'ont fait l'année même, tandis qu'en 2005, une partie a déjà liquidé l'année passée. On utilise les données de l'EIR 2008 pour attribuer aléatoirement une année de liquidation aux individus ayant liquidé en départ anticipé pour carrière longue et âgés de 60 ans ou moins à la date d'observation. Ces imputations sont effectuées par caisse, sexe et année afin de répliquer les répartitions d'années de liquidation observées dans l'EIR 2008. Ce point est nécessaire car le calage est ensuite effectué sur le stock de retraités, mais aussi sur le flux de liquidants.

Par ailleurs, les départs à 55 ans, possibles uniquement dans le cadre du dispositif de départ anticipé pour handicap¹⁶, ne sont pas clonés : on ne peut pas les identifier précisément dans l'EIR 2004.

1.2.3. Les spécificités du décalage temporel pour l'année 2009.

Certaines évolutions réglementaires survenues en 2009 font que, pour cette année de référence, la création d'un pseudo-EIR ne peut pas se faire pas 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. La durée d'assurance tous régimes augmente progressivement jusqu'à 172 trimestres pour les générations nées après 1952. Les conditions sur la durée d'assurance cotisée sont aussi durcies. Les pensions des pseudo-individus liquidant en 2009 dans le cadre du dispositif de départ anticipé pour carrière longue et ne remplissant pas ces nouvelles conditions sont donc exclues d'ANCETRE.

On tient aussi compte du nouveau seuil de liquidation sous la forme d'un versement forfaitaire unique à l'IRCANTEC : il faut désormais liquider avec au moins 300 points pour obtenir une pension. En deçà la liquidation s'effectue sous la forme d'un VFU. Les versements correspondants sont donc exclus de la base de données individuelle d'ANCETRE.

¹⁶ Ils représentent 0,1 % des départs à la retraite au régime général en 2011 (Drees, 2011)

Enfin, on intègre aussi la condition d'âge pour obtenir une pension de réversion au régime général et dans les régimes alignés, qui est rétablie à 55 ans.

1.3. Évolution des pensions et prise en compte de l'effet Noria

Les montants de pension sont calculés au 31/12 de chaque année. Les individus ayant par exemple 60 ans dans l'EIR 2004 sont considérés comme étant représentatifs des individus ayant 60 ans en 2005. On doit donc recalculer leur pension pour qu'elle soit représentative de la pension des individus âgés de 60 ans en 2005.

En général, la pension moyenne des retraités augmente d'année en année. Cette augmentation moyenne ne reflète pas l'évolution de l'avantage principal perçu, individuellement, par chaque retraité. La hausse de la pension provient en partie des revalorisations annuelles des pensions appliquées par chaque caisse de retraite, mais elle 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 qui disposent en moyenne de pensions plus élevées, et par le décès de retraités plus âgés percevant des pensions plus faibles en moyenne que les nouveaux retraités.

Pour rendre compte de ce mécanisme, dans ANCETRE, l'évolution de la pension moyenne intervient par différents canaux. Premièrement, pour les régimes les plus importants¹⁷, 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 donc revalorisées suivant l'évolution des pensions moyennes $\overline{pens}_{a,s,N}$ par âge et par sexe pour ces régimes (sachant que les retraités ayant 60 ans en 2004 sont considérés comme représentatifs des retraités ayant 60 les années suivantes). Par exemple :

$$pens_{a,s,2005} = pens_{a,s,2004} \times \frac{\overline{pens}_{a,s,2005}}{\overline{pens}_{a,s,2004}} \quad pens_{a,s,2006} = pens_{a,s,2004} \times \frac{\overline{pens}_{a,s,2006}}{\overline{pens}_{a,s,2004}}$$

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 :

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

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

¹⁷ Entre 2004 et 2007 : Régimes de base : CNAV, MSA « salariés » et « non-salariés », RSI « artisans » et « commerçants » ; Régime complémentaire : AGIRC ; régimes intégrés : Fonction publique d'État et CNRACL. A partir de 2008 : Ensemble des caisses de retraite participant à l'EACR.

Pour les régimes de retraite absents de l'EACR, et pour les régimes de l'EACR pour lesquels on ne dispose pas de données ventilées par génération depuis 2004, les pensions sont revalorisées uniquement en fonction des revalorisations légales appliquées par chaque régime. Pour les pensions de certains régimes (notamment pour certaines pensions versées par la CNAVPL au titre des avantages sociaux vieillesse), l'information disponible ne permet pas de revaloriser précisément les pensions¹⁸. Pour ces régimes, on s'attend donc à sous-estimer au fil des ans la pension moyenne.

Rappelons que, pour tous les régimes participant à l'EACR, les pensions moyennes totales¹⁹ 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 évoquée ci-dessus. Pour les régimes les plus importants en termes d'effectifs, le calage est réalisé génération par génération. Ce n'est cependant pas le cas pour les régimes plus petits : pour ces régimes, le calage conduit alors à augmenter le poids des pensions les plus élevées, dont bénéficient généralement les générations les plus jeunes, relativement aux pensions les plus faibles qui concernent majoritairement les générations plus âgées.

2. Pondérations et calage sur marge

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), puis le calage sur marges est utilisé pour caler les agrégats (nombres 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.

2.1. Les pondérations initiales

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 Métropole²⁰ 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.

Dans ANCETRE, 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. Le poids initial P d'un individu i de sexe s , appartenant à la génération g ²¹ inscrit dans la section sec du RNIPP et présent dans l'échantillon ANCETRE s'écrit ainsi :

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

avec NB = nombre total d'individus de sexe s , appartenant à la génération g inscrit dans la section sec du RNIPP 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.

Toutes les générations n'étant pas représentées dans l'EIR, des regroupements de générations R sont effectués (Aubert et Ducoudré, 2011, annexe 1). Ainsi par exemple, les individus nés en 1930 et

¹⁸ Soit parce que les revalorisations dépendent de la date d'acquisition et/ou de liquidation des points, ou de certaines conditions non mesurables dans l'EIR (par exemple les revalorisations des petites pensions à la MSA « non salariés »), soit parce qu'on ne dispose pas de l'information sur les revalorisations appliquées.

¹⁹ Plus précisément, on cale l'avantage principal de droit direct, sauf pour l'AGIRC où on cale la pension moyenne totale.

²⁰ L'appellation est trompeuse en ce qui concerne les personnes nées dans les départements d'outre-mer : ces personnes sont bien incluses dans la « section métropole » du RNIPP.

²¹ Il s'agit de la génération à l'issue de l'étape de décalage temporel

présents dans l'EIR 2008 sont considérés représentatifs de l'ensemble des individus des générations 1929 et 1930, ainsi que de la moitié des effectifs de la génération née en 1931. Dans ces cas, la formule donnant le poids initial d'un individu i est légèrement différente :

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

Trois corrections à cette formule doivent être signalées.

Premièrement, pour les générations les plus âgées, le RNIPP contient une proportion importante de faux vivants, les décès n'ayant pas toujours été enregistrés dans le répertoire. Pour les générations nées avant 1906, on a donc appliqué une correction aux effectifs du RNIPP à partir du ratio entre les effectifs nés en 1906 et les effectifs nés antérieurement, observé dans le Bilan Démographique de l'Insee.

Deuxièmement, une correction spécifique pour la mortalité est réalisée en ce qui concerne les personnes nées à l'étranger dans le RNIPP. Cette correction vise à tenir compte du fait que la mortalité peut y être sous-estimée, puisque les décès des personnes ayant quitté la France ne sont pas systématiquement renseignés, y compris pour des générations « récentes ». En effet, si on observe la mortalité d'une année sur l'autre (c'est-à-dire, pour une extraction donnée du RNIPP, le pourcentage de personnes vivantes une année donnée qui ne le sont plus l'année suivante), les coefficients de mortalité sont plus faibles pour les nés à l'étranger (section hors métropole SHM) que pour les nés en France (section métropole – SM).

On peut considérer que ce différentiel de mortalité constaté est dû uniquement au fait que certains décès ne sont pas renseignés dans le RNIPP-SHM, la mortalité devant sinon être similaire entre nés en France et nés à l'étranger à âge et sexe donné. Pour les nés à l'étranger, on « corrige » donc l'effectif des vivants chaque année en appliquant un coefficient égal au différentiel moyen de mortalité (par rapport aux nés en France) cumulé entre l'âge de 20 ans et l'âge en N^{22} . Par exemple, pour les personnes qui ont 40 ans en 2005 (génération 1965) :

$$Effectif(corrigé)_{1965}^{nés\ étranger} = Effectif(RNIPP - SHM)_{1965}^{nés\ étranger} \times Txcormort$$

avec

$$Txcormort = \frac{1}{4} \sum_{N=2006}^{2009} \frac{Txmort_{20ans}^{SM,N} * Txmort_{21ans}^{SM,N} * \dots * Txmort_{40ans}^{SM,N}}{Txmort_{20ans}^{SHM,N} * Txmort_{21ans}^{SHM,N} * \dots * Txmort_{40ans}^{SHM,N}}$$

où $Txmort_{20ans}^{SHM,2006}$ représente (par exemple) la mortalité à 20 ans constatée entre 2005 et 2006 pour les personnes nées hors de France (SHM) et ayant 20 ans en 2006. Le quotient de correction est bien toujours inférieur à 1 puisque, à chaque âge, la mortalité « constatée » dans le RNIPP est plus forte pour la section métropole que pour la section hors métropole $Txmort_a^{SM,N} < Txmort_a^{SHM,N} \forall$ âge a .

Enfin, une ultime correction est appliquée aux pondérations initiales des nés à l'étranger. Celles-ci sont réduites de 6,4% pour tenir compte de la proportion des nés à l'étranger dont la date de naissance n'est pas ou n'est que partiellement renseignée dans le RNIPP, et dont on suppose qu'ils ne deviennent jamais retraités. Dans la section hors métropole du RNIPP, la date de naissance est en effet incomplète pour une partie substantielle du répertoire (absence du jour, voire du mois de naissance). Les personnes qui ont une date de naissance incomplète peuvent correspondre à des personnes ayant séjourné en France pendant une durée plus courte : ces personnes ont donc des caractéristiques de carrière et de retraite différentes des personnes dont la date de naissance est

²² Le différentiel moyen de mortalité par âge est calculé sur les années 2005 à 2009, à partir d'une extraction du RNIPP au 31 décembre 2009. Il est calculé séparément pour les hommes et pour les femmes.

parfaitement connue dans le répertoire²³. L'échantillonnage de l'EIR se faisant sur le jour et le mois de naissance, on ne sélectionne par construction que des personnes dans cette dernière situation, et il ne s'agit donc pas d'un tirage totalement « aléatoire » dans la section hors métropole du RNIPP.

2.2. Le calage sur marges

2.2.1. Les données de calage : l'enquête annuelle auprès des caisses de retraite

L'EACR collecte depuis 2004 des données de cadrage auprès des principales caisses de retraite (CNAV, MSA, RSI, SRE, CNRACL, ARRCO, AGIRC, IRCANTEC, RATP, SNCF et CNIEG). L'enquête couvre approximativement 96% des pensions de droit direct versées par l'ensemble des caisses de retraite du système de retraite français²⁴. Elle collecte principalement des effectifs totaux de pensionnés en distinguant les bénéficiaires d'un droit direct et/ou dérivé au 31/12/N, les effectifs de nouveaux bénéficiaires d'une pension au cours de l'année N et les montants moyens d'avantage principal de droit direct et/ou dérivé correspondants pour chacune des caisses de retraite.

Des demandes complémentaires à l'enquête ont aussi 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 génération et pour certaines caisses par lieu de naissance²⁵. La collecte de données sur les pensions de réversion a aussi été progressivement améliorée : une collecte de tableaux ventilés par générations a été ajoutée à partir de 2008 pour la CNAV et le SRE, et étendue progressivement aux autres caisses en 2009. Les ventilations par générations ont aussi été étendues pour l'ensemble des tableaux (droits directs et dérivés) de 20 ans à 95 ans à partir de 2009 contre 50 à 85 ans auparavant.

2.2.2. Autres sources de données

Afin de prendre en compte les tendances d'évolution d'effectifs de retraités des caisses de retraite absentes de l'EACR, on a utilisé les données publiées chaque année dans le rapport de la Commission des Comptes de la Sécurité Sociale. Celles-ci portent sur les pensions de droit direct et dérivé versées par les régimes au 1^{er} juillet de chaque année. On estime des effectifs au 31 décembre de l'année N en calculant la demi-somme des effectifs au 1^{er} juillet des années N et N+1. Par contre, on ne dispose pas d'information sur les montants moyens de pension et sur les effectifs de nouveaux bénéficiaires.

2.2.3. Précisions sur 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 (cf. 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, d'où une augmentation significative du nombre de marges de calage (cf. Tableau 1).

Pour la mise en œuvre du calage sur marges, les marges contenant un faible effectif (inférieur à 2000 retraités) n'ont pas été conservées, afin de faciliter le calage. On a utilisé la méthode linéaire tronquée (cf. Sautory, 1993). Les bornes 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

²³ Plus précisément, il s'agit vraisemblablement de personnes qui ne sont restées que peu de temps en France, et qui ont donc une probabilité nettement plus faible que les autres personnes nées à l'étranger d'avoir acquis suffisamment de droits pour pouvoir liquider une pension de retraite versée en rente. La réduction des pondérations initiales des nés à l'étranger de 6,4 % correspond à l'hypothèse « extrême » qu'aucune des personnes dont le mois de naissance est inconnu dans le RNIPP n'a acquis de droits suffisants pour une liquidation en rente, et donc qu'aucune ne devient retraitée.

²⁴ Des travaux sont à l'étude pour élargir l'enquête à de nouvelles caisses de retraite.

²⁵ Selon une nomenclature correspondant à la scission du RNIPP en deux sections (nés en métropole et DOM / nés dans les TOM et à l'étranger).

rendre compte parfaitement de la montée des départs anticipés pour carrières longues. Cette contrainte est renforcée par le calage simultané de la pension moyenne par générations.

Tableau 1
Précisions sur le calage sur marge

| année | nombre de marges | ratio de poids minimum | ratio de poids maximum | poids minimum | poids maximum |
|-------|------------------|------------------------|------------------------|---------------|---------------|
| 2004 | 1003 | 0,1 | 5,79 | 3 | 1137,9 |
| 2005 | 1078 | 0,1 | 11,36 | 3,2 | 817,2 |
| 2006 | 1091 | 0,1 | 12 | 3,2 | 1034,7 |
| 2007 | 1085 | 0,1 | 14 | 3,2 | 1289,1 |
| 2008 | 1748 | 0,2 | 6,8 | 3,1 | 894,6 |
| 2009 | 2122 | 0,2 | 7,1 | 3,3 | 1119,5 |

2.2.4. Le calage pour les pensions de réversion

Concernant la réversion, le calage a été effectué depuis 2004, excepté pour l'AGIRC-ARRCO : les séries historiques ne permettent pas de remonter aussi loin pour l'ARRCO.

Par ailleurs, l'amélioration de la collecte des données pour ces caisses dans l'EIR 2008 crée une rupture par rapport à l'EIR 2004, qu'il n'a pas été possible de corriger. Une partie des pensions de réversion versées par l'ARCCO ont en effet été imputées dans l'EIR 2004, sur la base de données de cadrage issues de l'ancien système d'information de l'ARRCO. Or ce système a fortement évolué, et ne permet pas de reconstitution des séries d'effectifs corrigées jusqu'en 2004. A titre d'exemple, les effectifs de bénéficiaires d'un droit dérivé seul ont été révisés de -22% pour l'année 2006 entre les données de l'EACR pour cette année là et les données issues du nouveau système d'information.

In fine, ANCETRE ne permet pas actuellement de calculer de façon satisfaisante des effectifs de bénéficiaires d'une pension de réversion tous régimes confondus avant 2008.

3. Un complément : rétopolation des indicateurs tous régimes avant 2004

Dans ce qui a été décrit jusqu'ici, les pseudo-EIR construits par le modèle ANCETRE le sont à partir de versions de l'EIR qui leur sont antérieures chronologiquement. Les indicateurs issus du modèle ANCETRE sont donc disponibles pour les années à partir de 2004, les bases de données correspondantes étant construites à partir des EIR de 2004 ou de 2008.

Malheureusement, il n'est pas possible d'appliquer le modèle –avec la méthodologie décrite jusqu'ici– pour construire des bases sur des années de référence plus anciennes. Les vagues de l'EIR antérieures à 2004 (vagues 1988, 1993, 1997 et 2001²⁶) sont en effet sur le champ restreint des retraités nés en France, et manquent donc une proportion non négligeable des retraités.

La disponibilité de séries temporelles les plus longues possibles présente cependant un intérêt indubitable. Il est donc pertinent de chercher à développer une méthodologie spécifique pour rétopoler les séries avant 2004. Nous décrivons ci-dessous une méthode exploratoire pour réaliser cette rétopolation –« exploratoire » au sens où elle ne peut pas être appliquée en totalité à ce jour, faute de données de cadrage de l'EACR avant 2004.

Il s'agit plus précisément de construire des pseudo-EIR non pas à partir du dernier EIR disponible (puisque celui-ci ne couvre qu'une partie du champ des retraités pour les années avant 2004), mais à partir du *prochain* EIR disponible, c'est-à-dire le plus proche *après* l'année de référence étudiée. Par exemple, pour les années 2001 à 2003, la base de donnée individuelle réplique les données de l'EIR 2004, et non de celles de l'EIR 2001.

²⁶ Les vagues de l'EIR ont une périodicité de 4 ans. Les vagues 1993, 1997 et 2001 correspondent bien aux retraités vivant au 31 décembre de 1992, 1996 et 2000 respectivement, mais les montants de pension renseignés sont ceux des pensions versées au cours du premier trimestre suivant, ce qui explique la dénomination « décalée » d'une année.

La méthode se fonde sur l'utilisation des probabilités « individuelles » de décès à chaque âge, estimées selon la méthodologie de Aubert (2011), annexe 1. Pour chaque retraité présent dans l'EIR 2004 ou l'EIR 2008, il est en effet possible d'estimer une probabilité de décès entre une année N (antérieure à 2004) et l'année de référence de l'EIR (2004 ou 2008). Cette probabilité est conditionnelle à un certain nombre de caractéristiques individuelles du retraité : sexe, âge, mais aussi montant de pension tout régime, âge à la liquidation, fait d'être invalide ou ex-invalide, etc²⁷.

Par rapport à la méthode décrite dans les sous-parties précédentes, il ne s'agit pas de répliquer les individus pour représenter ceux des générations directement suivantes (en conservant l'âge à la date d'observation), mais de conserver les individus de la même génération en les rajeunissant (en diminuant leur âge à la date d'observation). Par exemple, alors que les retraités de 62 ans en 2006 (génération née en 1944) étaient « construits » à partir des retraités de la génération 1942 dans l'EIR de 2004 (62 ans à cette date), les retraités de 62 ans en 2002 (génération née en 1940) sont construits en rajeunissant les retraités qui ont 64 ans en 2004 (et appartiennent donc à la même génération 1940). En d'autres termes, il n'y a pas ici d'étape de « décalage temporel ».

Pour des raisons techniques, il est cependant nécessaire de disposer de toutes les générations après 66 ans, et pas seulement d'une génération sur deux. Les retraités des générations manquantes ayant entre 66 et 84 ans à la date d'observation sont donc créés par la méthode de clonage, décrite précédemment.

Au final, les étapes de la méthode sont les suivantes, pour construire la base ANCETRE « rétropolée » d'une année de référence N :

- 1) Clonage des retraités de générations impaires, ayant entre 66 et 84 ans à la date d'observation.
- 2) Modification des pondérations, en multipliant les poids individuels dans l'EIR (pondérations finales) par l'inverse de la probabilité (conditionnelles aux caractéristiques de l'individu) de survie entre l'année N et l'année de référence de l'EIR.
- 3) Correction du niveau des pensions en les divisant par les coefficients de revalorisation appliqués par le régime entre le premier janvier de l'année N+1 et la date de référence de l'EIR.
- 4) Suppression de la base de toutes les pensions (de droit direct ou dérivé) liquidées à partir du 1^{er} janvier de l'année N+1.
- 5) Calage sur marge sur les données agrégées de l'EACR.

Comme on l'a déjà signalé, l'étape 5) ne peut pour l'instant pas être réalisée, les données de l'EACR n'étant disponibles que depuis 2004. Cet obstacle n'est toutefois pas insurmontable, dans la mesure où la plupart des régimes peuvent réaliser un travail rétrospectif et fournir des données historiques au niveau agrégé. La méthode pourra donc être appliquée en totalité à l'avenir. Par ailleurs, les probabilités utilisées au cours de l'étape 2) sont des probabilités estimées. La méthode peut toujours être revue et améliorée, ce qui peut bien évidemment entraîner une modification à la marge des résultats. Enfin, une faiblesse de la méthode provient de l'étape 3) : on ne corrige en effet que des revalorisations légales appliquées à l'ensemble des retraités, et non d'éventuelles revalorisations spécifiques pour certaines catégories de retraités. L'écart entre les deux est un facteur de biais sur le montant moyen de pension, biais qui est en théorie corrigé par la procédure de calage sur marge lors de l'étape 5).

4. 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.

4.1. Les effectifs de retraités : avant et après calage

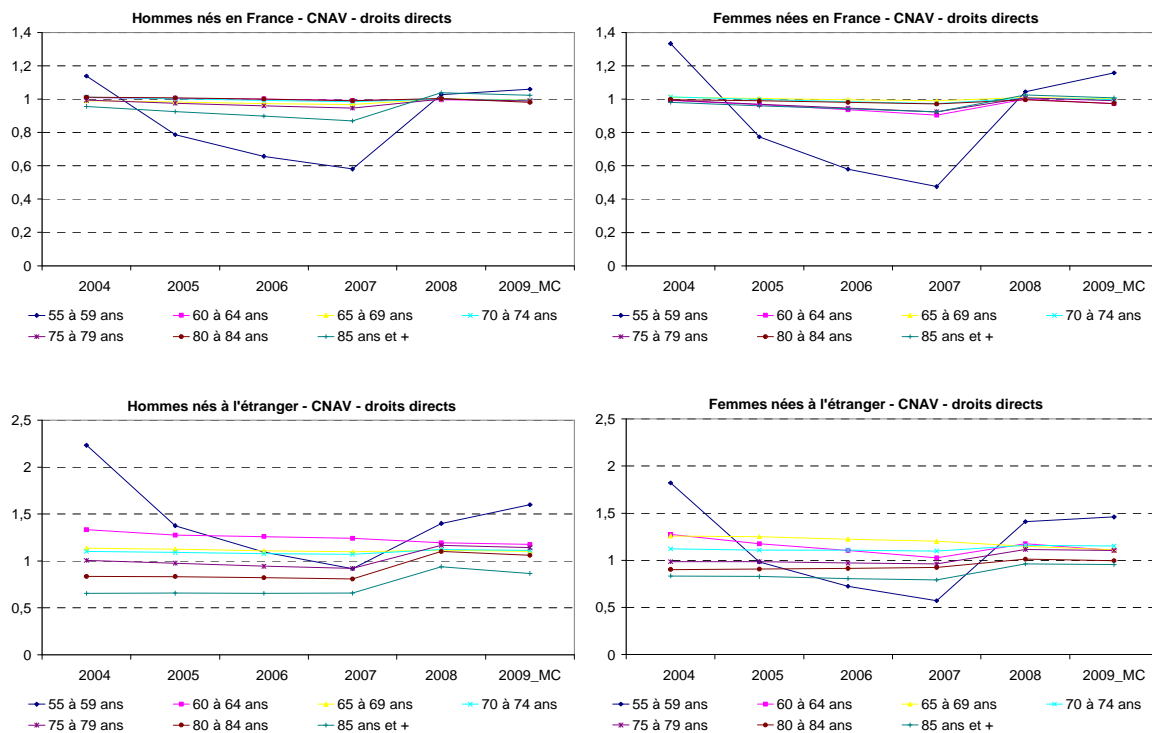
²⁷ On se reportera à Aubert et Christel-Andrieux (2010) pour une présentation complète de cette estimation. La méthode a ici été revue à la marge, notamment afin d'estimer des probabilités de décès annuelles et non quadriennales (à l'instar de Aubert, 2011). La spécification (choix des variables explicatives) a également été revue à la marge, mais les changements ne modifient pas les principes de la démarche empirique réalisée.

Si 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 (cf. graphiques 2 et 3).

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 courtes²⁸.

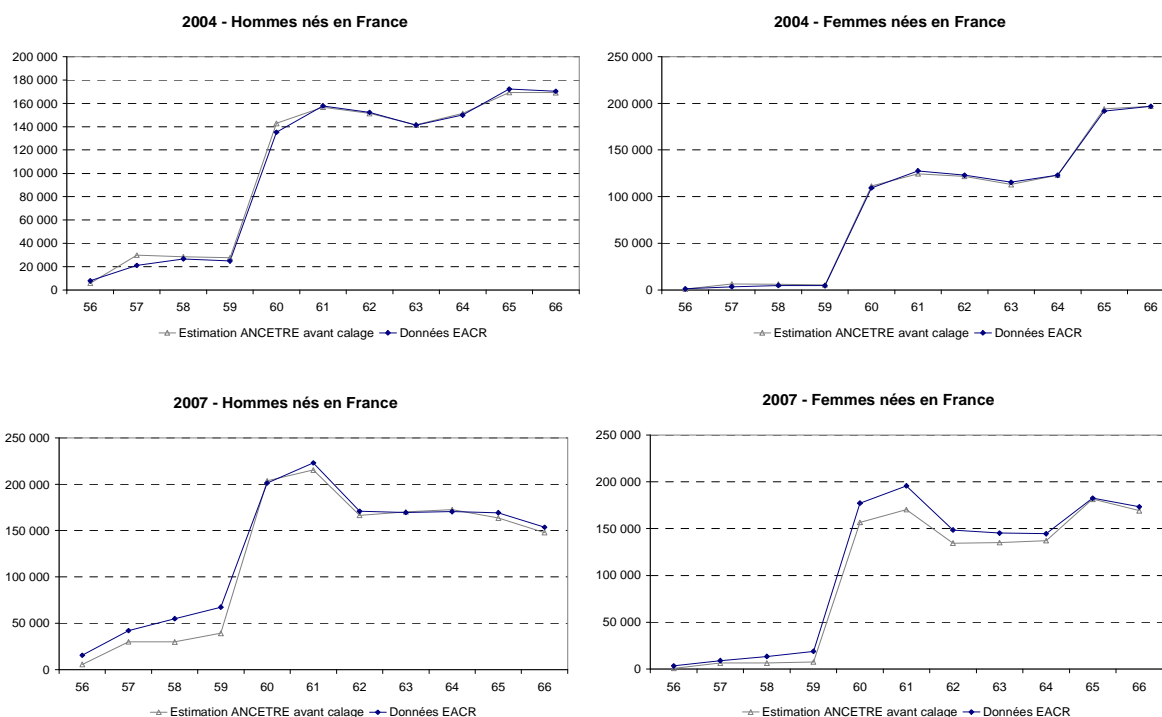
Graphique 1
Ratio entre les effectifs de retraités de droit direct calculés avec ANCETRE (pondérations initiales) et les effectifs tirés de l'EACR pour la CNAV



Les départs 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.

²⁸ On peut 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 nés à l'étranger à partir de 2006.

Graphique 2
Comparaison entre les effectifs pondérés ANCETRE avant calage et les effectifs de la CNAV tirés de l'EACR, par âge entre 56 et 66 ans



4.2. Ancienne méthode de calcul des indicateurs « tous régimes » vs 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.

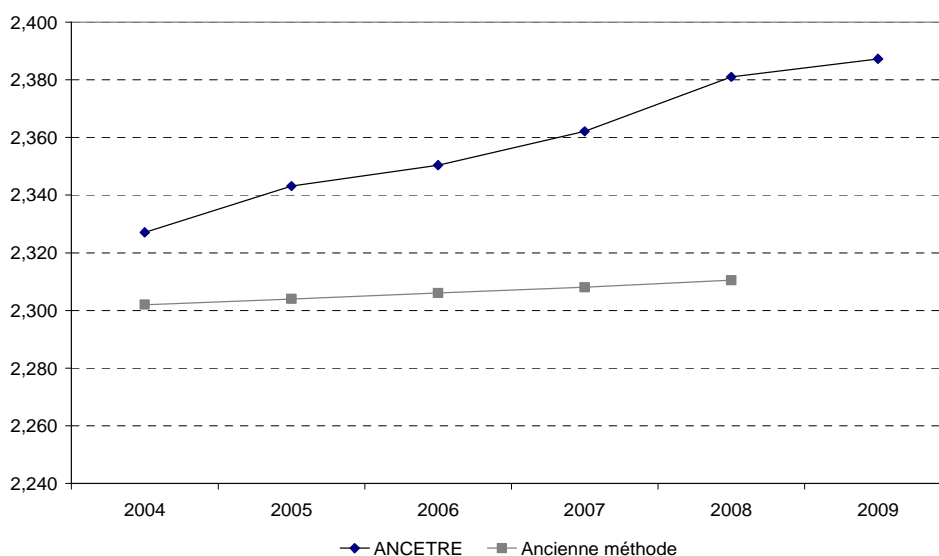
Par exemple, on estimait qu'un retraité percevait en moyenne 2,3 pensions de droit direct en 2007. Or les caisses de retraite servent 33,5 millions de pensions de droit direct cette même année. L'effectif total de retraités pour 2007 était donc estimé à 14,5 millions (33,5 divisé par 2,3).

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

Par exemple, le montant total des pensions de droit direct servies mensuellement par les caisses s'élevait à 16 milliards d'euros en 2007. Dès lors, l'estimation du montant moyen de pension de droit direct par retraité valait 1 108 euros mensuels pour 2007 (16 000 divisé par 14,5).

Ces calculs étaient effectués par la DREES en utilisant conjointement 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. Sa valeur était redressée, en prenant en compte les moins de 55 ans présents dans les régimes de la Fonction publique et dans les régimes spéciaux, considérés comme monopensionnés. Le taux de couverture des versements et des pensions par l'enquête annuelle (96 %) était également estimé d'après l'EIR. Les résultats (effectifs et montants) étaient corrigés en conséquence.

Graphique 3
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.

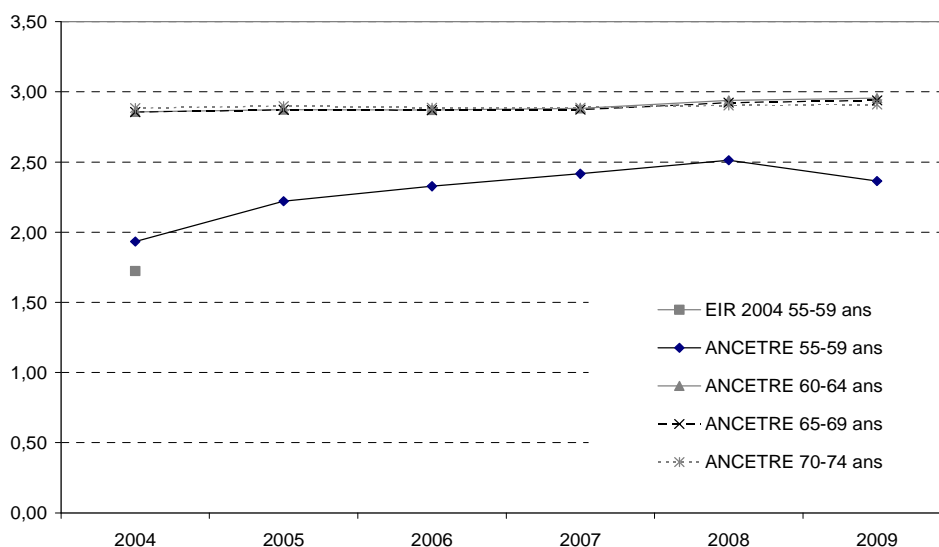
Cette méthode reposait fortement sur l'estimation du nombre moyen de pensions par retraité. Pour les années postérieures à 2004, une tendance *ad-hoc* était appliquée pour rendre compte de l'augmentation de ce nombre. Si on compare cette estimation avec celle obtenue par ANCETRE, on constate que la tendance²⁹ était largement sous-estimée entre 2004 et 2008 (cf. Graphique 3). 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 pensions³⁰ (cf. Graphique 4).

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 retraités 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).

²⁹ Le niveau estimé est également différent pour l'année 2004. L'écart s'explique par le traitement différent des pensions d'invalidité dans ANCETRE, ainsi que par la meilleure prise en compte des retraités qui sont hors du champ de l'EIR de 2004 (retraités de moins de 54 ans). Pour les retraités de 55 ans et plus, hors pensions d'invalidité, le nombre moyen de pensions par retraité est estimé à 2,3224 dans l'EIR 2004 contre 2,3509 dans ANCETRE pour la même année. Sur le même champ, la prise en compte des retraités de moins de 55 ans donne un nombre moyen de pensions par retraité de 2,302 avec l'ancienne méthode contre 2,329 avec ANCETRE (à nombre de pensions versées constant). Le traitement différent dans ANCETRE des pensions d'invalidité dans la Fonction publique et les régimes spéciaux explique l'écart résiduel.

³⁰ 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 (pension du régime de base, ainsi qu'une pension versée par l'ARRCO et éventuellement pension de l'AGIRC). A 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.

Graphique 4
Nombre moyen de pensions de droit direct par retraité selon la classe d'âge



Sources : DREES, ANCETRE, EACR, EIR 2004.

4.3. Basage sur l'EIR 2004 vs basage sur l'EIR 2008

Afin d'étudier la robustesse d'ANCETRE, on compare maintenant les résultats obtenus pour l'année 2008 à partir d'ANCETRE construit avec l'EIR 2004 (pseudo-EIR 2008 à partir de l'EIR 2004) et ceux obtenus avec l'EIR 2008 (qui, par construction, correspond rigoureusement au modèle ANCETRE pour l'année 2008). Cela permet d'analyser dans quelle mesure les indicateurs se déforment au fil du temps avec le calage.

Si 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 (cf. Tableau 2). Ce sont principalement les effectifs de femmes qui sont surestimés (+0,6% contre +0,2% pour les hommes ; cf. Graphique 5a). 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. Il s'ensuit une sous-estimation de l'avantage principal de droit direct moyen de 0,1% pour les hommes et de 0,4% pour les femmes en 2008 (cf. Graphique 5b).

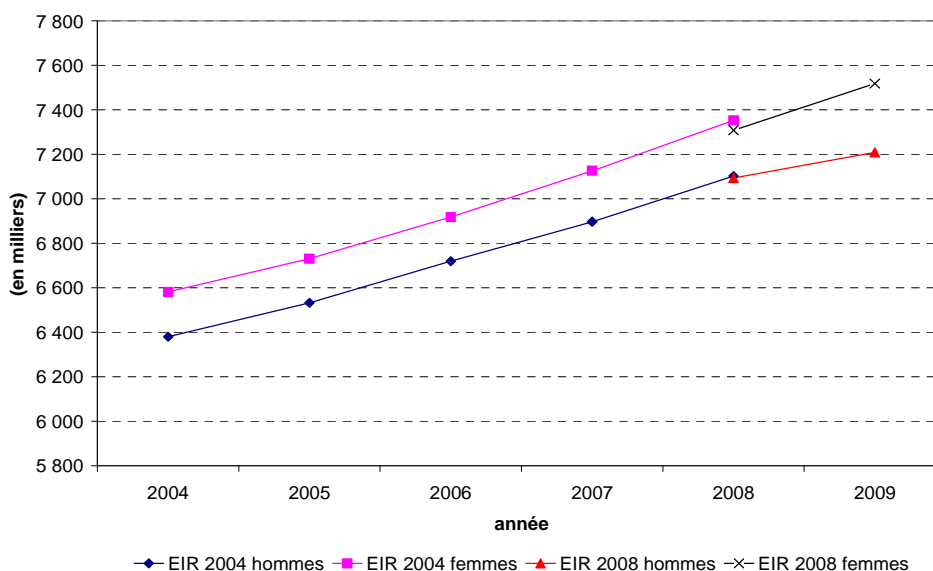
Ces écarts peuvent provenir de plusieurs facteurs, qui ne se cumulent pas nécessairement. Ainsi, la structure de polypension (à âge donné) pourrait se déformer au fil des générations. Dans ce cas, la prise en compte des évolutions démographiques, par le biais des pondérations initiales calculées par 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 (cf. Graphique 6). Les pensions moyennes par âge sont aussi globalement proches entre les deux méthodes (cf. Graphique 7 et Tableau 3). Si on regarde plus précisément pour les retraités âgés de 66 ans en 2008, la pension moyenne est légèrement surestimée pour les hommes (+1%) et sous-estimée pour les femmes (-1%) à partir d'ANCETRE construit avec l'EIR 2004 (cf. Tableau 4).

Finalement, la comparaison des deux méthodes d'estimation pour 2008 suggère que le « vieillissement » des EIR au fil des ans avec le modèle ANCETRE afin d'estimer les effectifs de retraités et les montants moyens de pension n'entraîne pas de dérive trop importante des indicateurs estimés. En particulier, l'écart observé entre les deux méthodes pour l'effectif de retraités en 2008

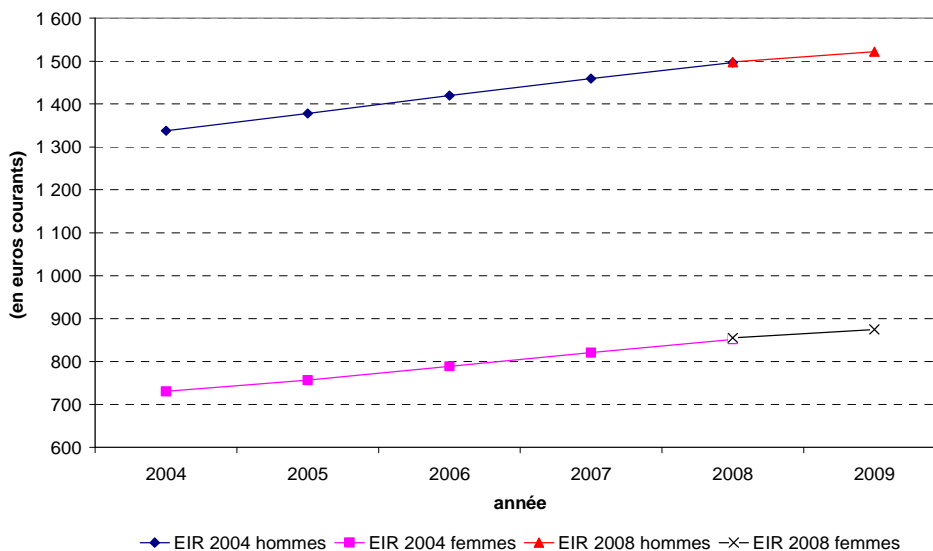
reste de faible ampleur, par comparaison aux variations annuelles du nombre de retraités (de 2007 à 2008 et de 2008 à 2009). Par ailleurs, un calage sur marges plus précis en 2008 puis pour les années 2009 et au-delà devrait permettre de réduire la dérive des indicateurs tous régimes en attendant le prochain EIR 2012.

Graphique 5a
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 5b
Pension mensuelle de droit direct moyenne tous régimes confondus, par sexe,
selon la vague de l'EIR retenue dans le modèle ANCETRE



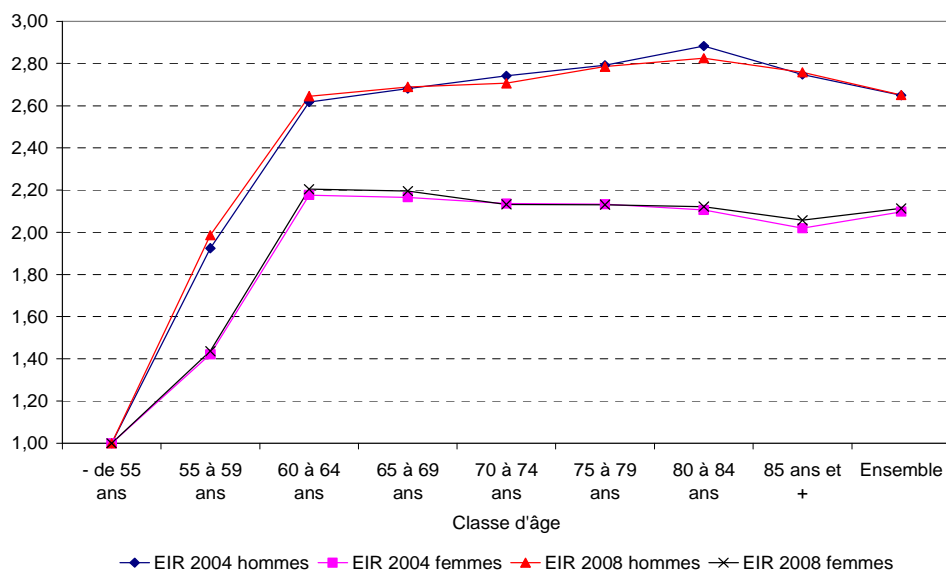
Note : la série « EIR 2004 » correspond aux pseudo-EIR annuels successifs construits à partir de cette vague de l'EIR, et la série « EIR 2008 » aux pseudo-EIR construits à partir de l'EIR de 2008.
 Source : DREES, ANCETRE.

Tableau 2
Contributions à l'écart des effectifs de retraité tous régimes
Par sexe et tranche d'âge en 2008 (réf. EIR 2004 / réf. EIR 2008)

| | Hommes | Femmes | Ensemble |
|-------------|--------|--------|----------|
| - de 55 ans | -0,07% | -0,06% | -0,07% |
| 55 à 59 ans | 0,37% | 0,11% | 0,24% |
| 60 à 64 ans | 0,35% | 0,16% | 0,25% |
| 65 à 69 ans | 0,26% | 0,19% | 0,23% |
| 70 à 74 ans | -0,32% | 0,17% | -0,07% |
| 75 à 79 ans | 0,14% | 0,18% | 0,16% |
| 80 à 84 ans | -0,28% | 0,04% | -0,12% |
| 85 ans et + | -0,31% | -0,20% | -0,26% |
| Ensemble | 0,13% | 0,60% | 0,37% |

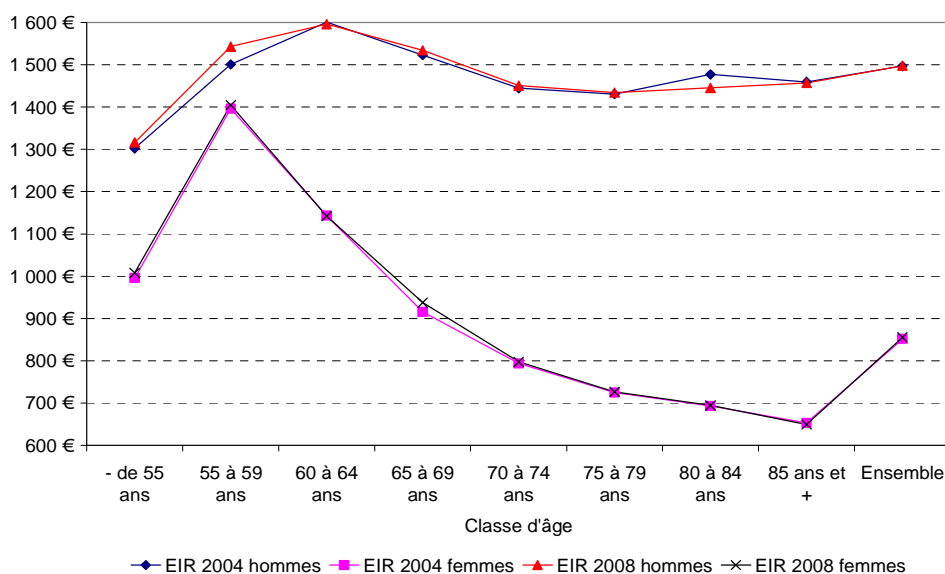
Source : DREES, ANCETRE.

Graphique 6
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 7
Pension moyenne (avantage principal de droit direct) tous régimes
Par sexe et tranche d'âge en 2008



Source : DREES, ANCETRE.

Tableau 3
Écart de pension moyenne (avantage principal de droit direct) tous régimes
Par sexe et tranche d'âge en 2008 (réf. EIR 2004 / réf. EIR 2008)

| | Hommes | Femmes | Ensemble |
|-------------|--------|--------|----------|
| - de 55 ans | -1,07% | -1,21% | -1,04% |
| 55 à 59 ans | -2,72% | -0,59% | -2,05% |
| 60 à 64 ans | 0,29% | 0,08% | 0,27% |
| 65 à 69 ans | -0,72% | -2,37% | -1,29% |
| 70 à 74 ans | -0,41% | -0,44% | -0,81% |
| 75 à 79 ans | -0,25% | -0,23% | -0,27% |
| 80 à 84 ans | 2,21% | -0,24% | 0,70% |
| 85 ans et + | 0,17% | 0,57% | -0,21% |
| Ensemble | -0,07% | -0,43% | -0,27% |

Source : DREES, ANCETRE.

Tableau 4
Écart de pension mensuelle moyenne tous régimes
estimée pour la génération née en 1942 à 66 ans en 2008

| | Hommes | Femmes | Ensemble |
|--|--------|--------|----------|
| Pseudo-EIR 2008 construit à partir de l'EIR 2004 | 1570 | 955 | 1273 |
| EIR 2008 | 1556 | 964 | 1270 |
| Écart (2004/2008) | 0,87% | -0,86% | 0,18% |

Source : DREES, ANCETRE.

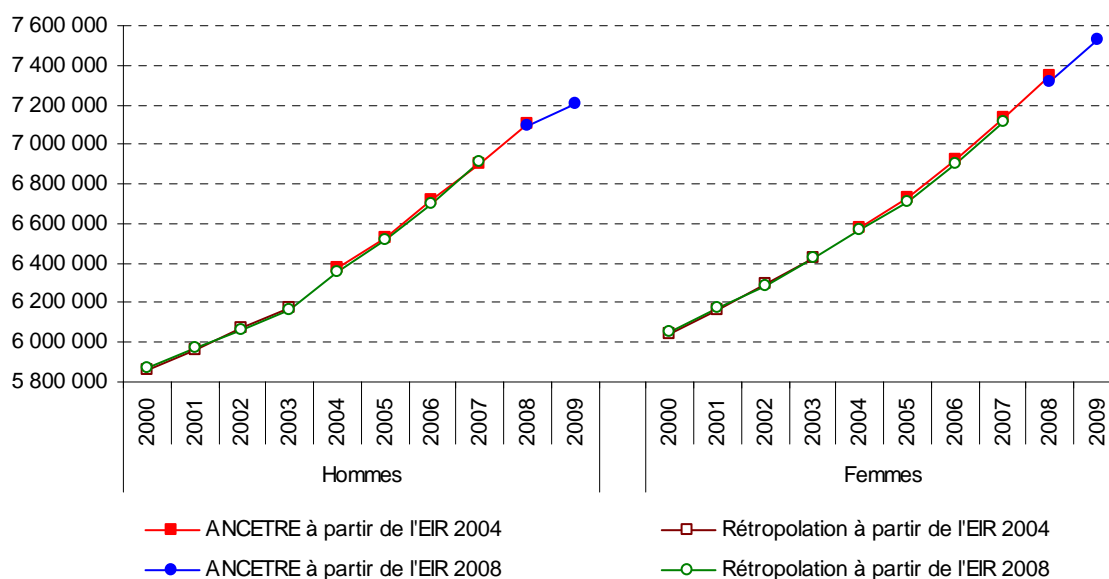
4.4. Comparaison des données projetées et des données rétropolées pour les années 2000 à 2003

La méthode de construction de bases ANCETRE rétropolées (cf. *supra*) peut être appliquée aux données de l'EIR 2004 ou bien à celles de l'EIR 2008. Néanmoins, les estimations sont d'autant plus imprécises que l'on s'éloigne de l'année de référence de l'EIR à partir desquelles la base ANCETRE est construite. Nous ne tentons donc pas de rétropolation avant 2000. Par ailleurs, on rappelle que, à ce jour, aucune donnée de calage n'est disponible pour les années avant 2004. Nous n'appliquons donc ici que les étapes 1) à 4) de la méthode décrite dans la troisième partie de cette étude.

De la même façon que la comparaison à l'EIR 2008 de la base ANCETRE 2008 construite à partir de l'EIR 2004 permettait de vérifier l'adéquation du modèle aux données réelles (cf. supra), on peut ici confronter les résultats de la base ANCETRE 2004 construite à partir de l'EIR 2008 aux données de l'EIR 2004. La comparaison, pour les années 2000 à 2003, des bases ANCETRE issues de l'un ou l'autre des deux EIR permet par ailleurs d'apprécier la robustesse des résultats.

Ces comparaisons sont globalement satisfaisantes. Pour le nombre total de retraités de droit direct, les deux estimations pour 2004 sont par exemple très proches, et l'écart est de faible ampleur par comparaison avec les variations annuelles de l'indicateur (Graphique 8). Les écarts sont relativement plus importants pour d'autres indicateurs, mais restent d'ampleur raisonnable.

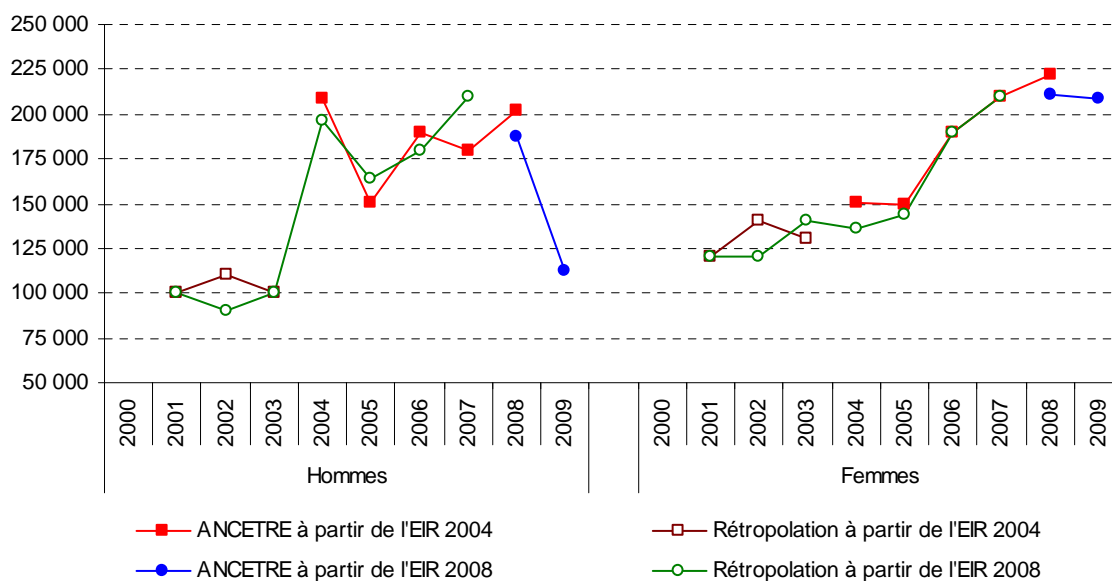
Graphique 8
Nombre de retraités de droit direct tous régimes confondus, par sexe
selon l'EIR de référence utilisé pour construire les bases



Champ : tous retraités de droit direct, nés et résidents en France ou à l'étranger

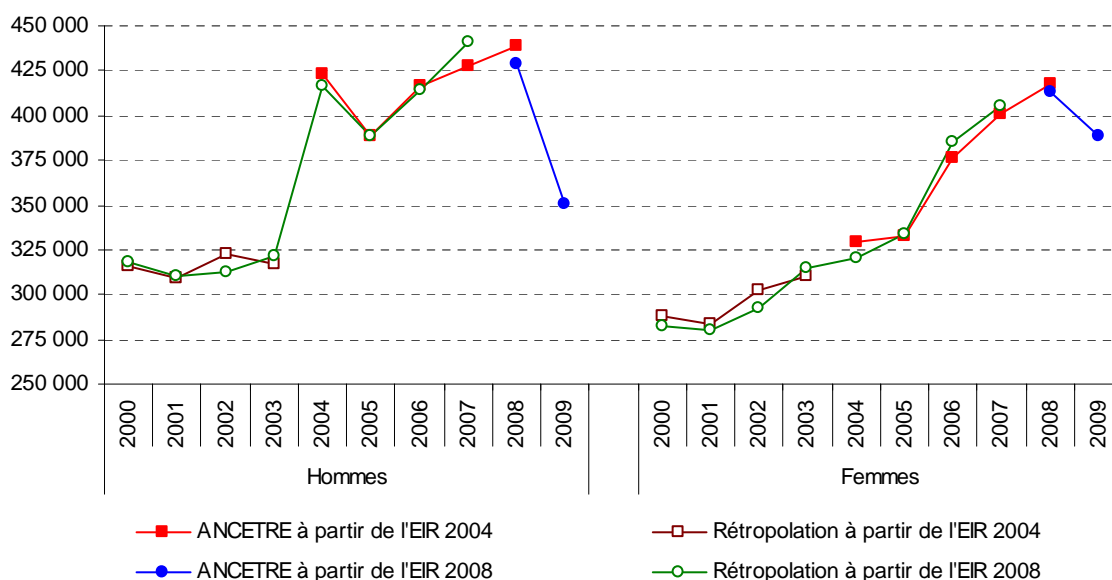
Source : DREES, ANCETRE

Graphique 9
Variation annuelle du nombre de retraités de droit direct tous régimes, par sexe
selon l'EIR de référence utilisé pour construire les bases



Champ : tous retraités de droit direct, nés et résidents en France ou à l'étranger
 Source : DREES, ANCETRE

Graphique 10
Nombre de primo-liquidants tous régimes
(nouveaux retraités d'un premier droit direct au cours de l'année), par sexe
selon l'EIR de référence utilisé pour construire les bases



Champ : retraités liquidant un premier droit direct au cours de l'année, nés et résidents en France ou à l'étranger
 Source : DREES, ANCETRE

5. Limites et pistes d'amélioration du modèle

L'élargissement envisagé du champ de l'EACR aux caisses ne participant pas actuellement à l'enquête permettra 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. Cette amélioration concerne en premier lieu les vagues à venir du modèle (à partir d'ANCETRE 2010), mais elle pourra également concerner éventuellement certaines vagues passées, grâce à la fourniture, par certaines caisses participantes ou entrantes dans le champ de l'EACR, de données historiques au format de cette enquête.

ANCETRE reste adapté essentiellement pour fournir des indicateurs qui sont issus des variables pour lesquelles il y a 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 également d'autres indicateurs tous régimes (proportions associées à certains dispositifs tels que la décote, surcote, le minimum contributif, etc.). Mais ces indicateurs seraient alors plus fragiles, du fait de l'absence de calage spécifique sur les variables correspondantes. En particulier, la cohérence régime par régime avec des données agrégées annuelles ne serait pas, pour ces variables, assurée. 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. Les prises en compte des effets de certaines réformes telles que le plafonnement de certains dispositifs en fonction de l'ensemble des revenus n'est pas non plus envisageable actuellement lors du clonage, les informations disponibles dans les EIR n'étant pas suffisantes pour ce type de réformes. 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 des réformes de ce type 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. A 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 en 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.

Bibliographie

- [1] Aubert P. et V. Christel-Andrieux, (2010) "La mortalité différentielle des retraités : estimation à partir de l'échantillon interrégimes de retraités et applications", DREES Documents de travail – série 'Études et Recherches', n°100, Juillet.
- [2] Aubert P. et B. Ducoudré (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.
- [3] Drees, « Les retraités et les retraites en 2009 », Collection Études et Statistiques, juin 2011.
- [4] 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.