

La modélisation des trajectoires professionnelles dans le modèle Prisme

Caroline Berteau-Rapin *, Paul Beurnier * et Émilie Denayrolles *

Le modèle de microsimulation de la Cnav, Prisme (Projection des retraites individuelles : simulation, modélisation, évaluation), a été créé en 2004. Ce modèle répondait à un besoin d'évaluation des effets des réformes des retraites, tant du point de vue financier pour le régime, que du point de vue de l'assuré, à travers l'évolution de son montant de pension ou de son âge de liquidation. Les nombreuses non-linéarités induites par le mode de calcul de la retraite au Régime Général ont rendu nécessaires les évaluations basées sur un grand nombre de situations individuelles, ce qui est l'essence même de la microsimulation. Prisme repose sur les fichiers de gestion de la Cnav retraçant la totalité de la carrière des assurés au régime général et, de plus en plus, dans les autres régimes. Il s'appuie sur un échantillon au 1/20^e de près de 5 millions d'individus.

Le module « carrière » permet de modéliser les statuts d'activité de chaque individu pour chaque année et chaque trimestre projetés. Il permet ainsi de simuler le partage de la population entre actifs et inactifs, et au sein des actifs, entre salariés du régime général, salariés d'autres régimes (alignés et non alignés), chômeurs, salariés en arrêt maladie, invalides. Depuis 2009, ces situations sont modélisées par un enchaînement d'équations logistiques, estimées sur les transitions de carrières des années 2005 à 2007. Elles sont déclinées par genre, tranches d'âge glissantes et lieu de naissance. Un salaire annuel est ensuite attribué aux individus pour lesquels un report d'activité salarié au régime général est affecté.

Le module « carrière » permet donc de simuler de manière fine les trajectoires professionnelles des individus et d'estimer les conditions de leur départ (date de départ, durées d'assurance, montant de leur retraite, etc.). Des indicateurs de suivi des carrières projetées sont présentés dans ce papier et permettent ainsi de valider la cohérence de cette modélisation.

Rappel :

Les jugements et opinions exprimés par les auteurs n'engagent qu'eux mêmes, et non les institutions auxquelles ils appartiennent, ni a fortiori l'Insee.

Codes JEL : J21, J26, J60.

Mots clés : microsimulation, trajectoires professionnelles, projections, régime de retraite.

* Caisse nationale d'assurance vieillesse (Cnav).

La Cnav et son réseau régional, composant l'Assurance retraite, gèrent les droits à la retraite des salariés du secteur privé. Le régime de sécurité sociale de ces derniers, appelé régime général, est le premier régime de retraite en France. En 2013, il regroupait 68,7 % des actifs, 17,7 millions de cotisants, 13,5 millions de retraités et versait plus de 105 milliards d'euros de prestations¹. En 2001, 95 % des monocotisants et plus de 90 % des polycotisants ont cotisé au régime général² (Chaput, 2005).

Les évolutions démographiques en France, avec la stagnation du nombre de cotisants et l'envolée des effectifs de retraités, ont contraint les pouvoirs publics à adapter les dispositifs de retraite afin de réduire, sinon d'éviter, la dégradation de l'équilibre des régimes de retraite. Dans cet esprit, plusieurs réformes ont été engagées depuis 1993. Ces réformes, le plus souvent de nature paramétrique, ont accentué les effets individuels lors du passage à la retraite, rendant obsolètes les modélisations fondées

sur des individus moyens utilisées jusqu'alors (cf. encadré 1).

L'évolution du système de retraite a rendu nécessaire l'utilisation de modèles de microsimulation qui intègrent les comportements au niveau individuel. Il est ainsi possible de mesurer plus précisément les conséquences engendrées par l'introduction de nouvelles mesures réglementaires. De plus, la prise en compte des évolutions législatives et démographiques est plus aisée avec ce type de modèle. La microsimulation dynamique est utilisée en France et à l'étranger (Li, 2012) pour effectuer des prévisions de long terme et prévoir ainsi les dépenses de retraite.

1. www.lassuranceretraite.fr

2. Un monocotisant est une personne qui a cotisé exclusivement dans un seul régime au cours de sa carrière. Un polycotisant a cotisé dans plus d'un régime de base.

Encadré 1

PRINCIPALES RÉFORMES DES RETRAITES DU RÉGIME GÉNÉRAL DEPUIS 1993

Réforme 1993

- 1) Hausse de la durée de cotisation : de 150 à 160 trimestres à raison d'un trimestre par génération.
- 2) Élargissement de la période de référence du salaire annuel moyen pour le calcul de la pension : de 10 à 25 meilleures années.
- 3) Officialisation de l'indexation des pensions sur l'indice des prix à la consommation.

Réforme 2003

- 1) Hausse de la durée d'assurance en fonction de l'augmentation de l'espérance de vie à 60 ans et alignement de la durée de référence prise en compte pour le calcul des pensions sur la durée nécessaire pour bénéficier du taux plein.
- 2) À partir de 2009, hausse de la durée d'assurance d'un trimestre par an pour atteindre 164 trimestres en 2012.
- 3) Mise en place d'un dispositif pour les carrières longues permettant aux individus ayant commencé à travailler jeune de partir en retraite anticipée (avant l'âge légal).
- 4) Instauration d'un dispositif de surcote et allègement de la décote.

Réforme 2010

- 1) Relèvement progressif de l'âge légal : de 60 ans à 62 ans (à raison de 4 mois pour les individus nés entre le 1^{er} juillet 1951 et le 31 décembre 1951, puis 5 mois supplémentaires jusqu'à la génération 1955).
- 2) Relèvement progressif de l'âge au départ sans décote : de 65 à 67 ans (au même rythme que la hausse de l'âge légal).
- 3) Mise en place d'un dispositif de retraite anticipée pour pénibilité.

Décret du 2 juillet 2012

Élargissement du dispositif des retraites anticipées pour carrière longue aux individus justifiant d'un début d'activité avant 20 ans (départ à 60 ans).

Réforme 2014

- 1) Hausse de la durée d'assurance à raison d'un trimestre tous les 3 ans pour atteindre 172 trimestres pour la génération 1973.
- 2) Création du compte personnel de prévention de la pénibilité.
- 3) Revalorisation des pensions au 1^{er} octobre (au lieu du 1^{er} avril).
- 4) Acquisition d'un trimestre d'assurance retraite après avoir cotisé sur un revenu équivalent à 150h de Smic (contre 200h auparavant).

En France, le modèle Destinie (Blanchet *et al.*, 2010), développé depuis le milieu des années 1990 par l'Insee, a été le premier modèle de microsimulation dynamique à simuler les dépenses de retraite³. Il couvre l'ensemble du système de retraite. Destinie 2 est basé sur l'enquête *Patrimoine 2009* de l'Insee qui regroupe des informations sur les carrières et les liens familiaux. En projection, les transitions professionnelles s'appuient sur l'EIC⁴ (échantillon interrégimes de cotisants) 2005 pour les milieux et fins de carrière et l'enquête *Génération 1998* (Céreq) pour l'insertion dans la vie active. La méthodologie retenue pour simuler les transitions de carrière est celle des logits emboîtées, elles sont déclinées par statut de départ, genre et période dans la vie active (Bachelet *et al.*, 2014).

Deux autres modèles sont eux aussi des modèles généralistes visant à couvrir l'ensemble du système de retraite. L'Institut des politiques publiques (IPP) développe actuellement son propre modèle de microsimulation en partenariat avec l'Insee, le modèle PensIPP. Ce modèle, dont l'objectif est de projeter les retraites à long terme, est basé sur Destinie 2.

La Drees (Direction de la recherche, des études, de l'évaluation et des statistiques)⁵ a développé de son côté le modèle Trajectoire (Duc, 2013) qui simule la carrière des individus pour tous les régimes de retraite. Il s'appuie sur l'EIC 2009. Quatre-vingt-dix-huit états sont définis dans le modèle Trajectoire, il s'agit d'états de non-emploi et d'emploi déclinés par régimes (Cnav, MSA⁶ salarié, MSA exploitant, RSI⁷, Fonction publique, professions libérales). Deux états sont modélisés annuellement : un état principal et un état secondaire. Les probabilités de transitions de carrière sont estimées de manière à respecter celles observées dans l'EIC.

À côté de ces modèles généralistes, il existe ou sont en cours de développement des modèles spécifiques aux régimes les plus importants. Tel est le cas du modèle Prisme⁸ qui a été le premier de ce type et qui fait l'objet de cet article. Il a été développé par la DSPR (Direction statistiques, prospective et recherche) de la Cnav à partir de 2004 (Poubelle, 2006). L'objectif de ce modèle est double. Il effectue régulièrement des prévisions de long terme (à horizon 2060) lors des exercices menés par le COR (Conseil d'orientation des retraites). Prisme est aussi utilisé en tant qu'outil de prévision de court terme (4 ans) des dépenses du régime général pour la Commission des comptes de la Sécurité sociale

(deux fois par an). Prisme est le seul modèle, parmi ceux mentionnés précédemment, qui permet à la fois la réalisation de projections de long terme ainsi que des prévisions budgétaires à plus court terme utilisables en gestion. Il permet également de simuler les évolutions réglementaires. Il a été mobilisé afin d'évaluer les mesures constituant la réforme 2010. Prisme a également permis aux pouvoirs publics de définir les contours du décret du 2 juillet 2012 qui élargit les conditions d'accès à la retraite anticipée pour carrière longue.

Quatre modules composent le modèle Prisme. Le module « démographie » regroupe les procédures d'estimation des naissances, des décès et des migrations. Dans le module « carrière », les trajectoires professionnelles des individus sont projetées avec, en cas d'activité salariée, l'estimation du salaire. L'AVPF (allocation vieillesse des parents au foyer) est également estimée. La carrière des individus est projetée selon un pas trimestriel. Le module « départs en retraite » modélise les départs à la retraite (droits propres) et applique les formules de calcul de la pension au régime général. Enfin, une fois la projection des droits propres⁹ réalisée, le module « réversion » évalue pour chaque décès d'un assuré les droits à réversion¹⁰ du conjoint survivant. L'architecture du modèle Prisme est présentée en figure I.

Prisme n'a cessé d'évoluer depuis sa création. D'abord fondé sur un échantillon au 1/100^e, il repose depuis 2006 sur un échantillon au 1/20^e (soit 5 millions d'individus). En 2007, une mortalité différenciée par type de pension a été introduite dans le module « démographie ». Dans les premières versions, l'estimation des carrières était réalisée à l'aide d'une matrice de transition. Depuis 2009, les carrières sont estimées à partir d'un enchaînement d'équations logistiques. Le module « départs en retraite » s'appuyait également sur une matrice mais, depuis 2006, le choix de départ est modélisé par des équations logistiques. Autre amélioration

3. Il a notamment été sollicité lors de la réforme des retraites de 2003.

4. L'EIC 2005 est un échantillon conçu par la Drees dont le taux de sondage est d'une génération sur quatre (de 1934 à 1974) et pour chaque génération tirée, les individus nés les dix premiers jours du mois d'octobre, soit entre 20 000 et 25 000 individus par génération.

5. Ministère chargé des Affaires sociales et de la Santé.

6. Mutualité sociale agricole.

7. Régime social des indépendants.

8. *Projection des Retraites Individuelles : Simulation, Modélisation, Évaluation.*

9. Droits à pension au titre de l'activité salariée.

10. Droits à pension au titre de l'activité salariée du conjoint décédé.

importante, en 2010, la projection des départs en retraite est passée d'un pas trimestriel à un pas mensuel afin de s'adapter au rythme de la montée en charge de l'augmentation de l'âge de départ en retraite de la réforme des retraites (hausse de l'âge légal de quatre mois d'abord, puis de cinq mois). Plus récemment, le compte personnel de prévention de la pénibilité a été intégré dans Prisme.

L'objet de cet article est de présenter le module « carrière » de Prisme. Ce module joue un rôle central dans la projection des retraites et des masses financières, celles-ci découlant principalement des carrières professionnelles. La richesse des données sur lesquelles s'appuie le modèle Prisme permet une simulation très fine des trajectoires professionnelles en utilisant ces équations logistiques (cf. infra, modélisation des carrières).

Dans une première partie, nous présenterons les bases de données utilisées pour l'estimation

des trajectoires professionnelles. Dans une deuxième partie, nous décrirons la modélisation utilisée pour les transitions de carrière. Enfin, la troisième partie sera consacrée à l'étude des carrières projetées.

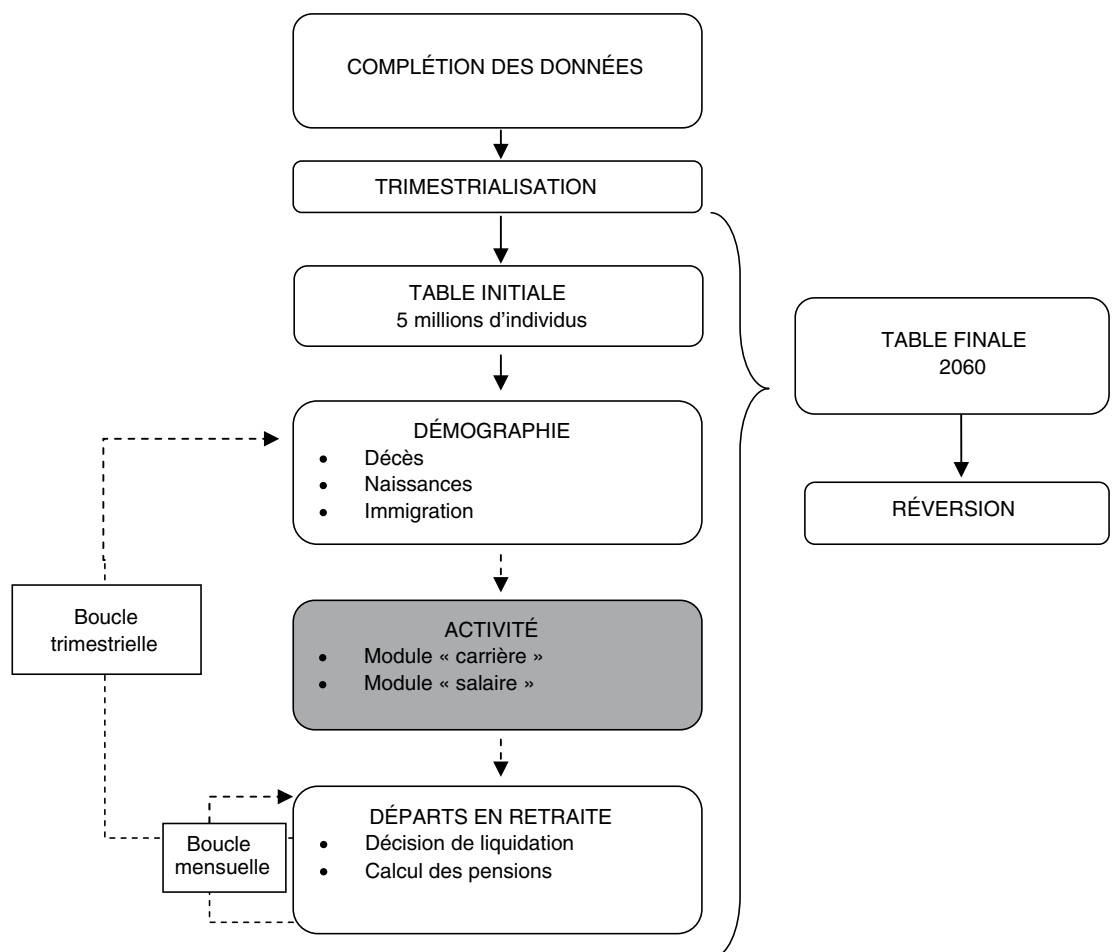
Présentation des données

Les données individuelles utilisées

La Cnav dispose de fichiers de gestion qui regroupent les éléments de carrière des salariés au régime général et qui permettent le paiement des retraites. Ces bases contiennent l'intégralité des personnes affiliées à la Sécurité sociale, soit plus de 90 millions d'individus¹¹.

11. Les effectifs de personnes affiliées à la Sécurité sociale sont supérieurs à la population totale française. Ces effectifs regroupent toute personne ayant cotisé au moins un trimestre, née ou non en France et présente ou non sur le territoire français.

Figure I
Architecture du modèle Prisme



Les données concernant l'activité professionnelle des salariés proviennent du SNGC (Système national de gestion des carrières). Il mémorise la totalité de la carrière des assurés sociaux pour le calcul de leur retraite : salaires cotisés, trimestres validés au régime général et dans les autres régimes, périodes de chômage, de maladie, etc. L'information sur les salaires provient des DADS (déclarations annuelles de données sociales), des DNT (déclaration nominative trimestrielle) et des Cesu (chèques emploi service universel). Les données relatives au chômage sont fournies par Pôle emploi, les données maladies sont transmises par la Cnamts (Caisse nationale d'assurance maladie des travailleurs salariés) et les cotisations AVPF par la Cnaf (Caisse nationale des allocations familiales).

Prisme s'appuie sur un échantillon au 1/20^e actualisé tous les deux ans et, depuis la création du modèle, sur une sélection invariante de 5 clés de contrôle (les deux derniers chiffres qui suivent le NIR¹²) prises initialement au hasard parmi les 97 existantes. Ce mode de tirage de l'échantillon permet de conserver le panel constitué entre deux extractions, enrichi des nouveaux assurés au régime général correspondant au critère de sélection (naissances et immigrants). Les données portant sur la retraite sont actualisées tous les ans (évolutions de pension, décès éventuels...) pour une meilleure précision des prévisions de court terme. Cet échantillon compte ainsi près de 5 millions d'individus.

Avant l'utilisation de ces données pour la projection, des procédures de complétions sont appliquées aux informations, éventuellement lacunaires, susceptibles d'avoir une incidence sur la retraite (périodes d'activité effectuées dans un autre régime et non encore connues au SNGC). De même, certaines variables, inexistantes dans les fichiers de gestion de la Cnav, mais essentielles pour l'objectif recherché, sont créées.

Complétion et création de données relatives aux carrières

Concernant le module « carrière » de Prisme, trois étapes préalables à la modélisation sont nécessaires. Il s'agit de la création d'une variable contenant l'âge de fin d'études, de la complétion de périodes validées dans d'autres régimes et de la trimestrialisation des reports d'activité au sein d'une année.

L'âge de fin d'études est une information importante pour des événements touchant

l'assuré en projection. Il influe notamment sur la fécondité, la croissance salariale et les transitions autour de l'emploi. Cette variable, absente des fichiers administratifs de la Cnav, est complétée à l'aide des salaires contenus dans les fichiers. Schématiquement, une forte hausse de salaire avant 30 ans est considérée comme un signal de fin d'études l'année précédant cet accroissement. Pour les individus dont les reports de salaires annuels ne connaissent pas une hausse jugée suffisante pour y reconnaître le signal, ou pour ceux qui n'ont pas de report de salaire avant 30 ans, un âge de fin d'études est affecté afin que la répartition globale respecte celle de l'ensemble des déclarants de l'enquête *Emploi* de l'Insee.

Les bases de données de la Cnav, si elles sont exhaustives quant au nombre d'individus qui y figurent, ne le sont pas toujours au regard des types de périodes renseignées dans la carrière. En effet, toute personne immatriculée par la Sécurité sociale apparaît dans ces fichiers, qu'elle ait déjà cotisé ou non au régime général. Cependant, à l'inverse des périodes concernant le régime général qui sont renseignées au fil de l'eau, la majorité des périodes relevant des autres régimes est, au mieux, renseignée pour les assurés atteignant un âge donné (par exemple, 53 ans pour la MSA), voire au moment des opérations du droit à l'information¹³ ou du départ en retraite de l'assuré (pour la fonction publique d'État ou certains régimes spéciaux, par exemple).

Aussi, **la complétion des périodes « autres régimes »** dans Prisme remplit trois objectifs principaux en partant des données de cadrage issues de l'EIC (échantillon interrégimes de cotisants), mis en place par la Drees et alimenté par l'ensemble des régimes de retraite :

- le respect d'une part de monocotisants (individus dont la carrière est exclusivement effectuée au régime général), les individus étant sélectionnés pour respecter les répartitions de durées validées au régime général constatées au sein des deux catégories créées (monocotisants/polycotisants)¹⁴ ;

12. Numéro d'inscription au répertoire, plus connu comme « numéro de sécurité sociale ».

13. Ces opérations ont pour objectif d'informer au plus tôt les assurés sur leurs droits à la retraite. Il est par conséquent nécessaire de disposer de tous les éléments de carrière. Ainsi, dès l'âge de 44 ans, si un individu constate une anomalie ou un oubli sur son relevé de carrière, il peut le signaler en effectuant une régularisation de carrière.

14. Données de calage : parts de monocotisants et de polycotisants et durées cotisées au régime général issues de l'EIC (extrapolation linéaire pour les générations non présentes dans l'EIC).

- le respect des distributions de durée tous régimes au sein de chacune de ces catégories (monocotisants/polycotisants)¹⁵ ;

- le respect d'une cible annuelle de cotisants « autres régimes » par type de report de 1999 à 2009¹⁶.

Une probabilité d'être polycotisant est estimée à l'aide de régressions logistiques estimées sur les carrières entièrement connues (celle des jeunes retraités). Des périodes « autres régimes » sont donc imputées et placées dans la carrière au niveau individuel, en respectant les données de calage déclinées par genre, génération et lieu de naissance (France ou étranger).

Parmi les périodes relevant du régime général, les données relatives au service militaire sont les seules à ne pas avoir été remplies en continu. Cependant, des échanges informatiques ayant été mis en place, certaines générations voient désormais leurs périodes militaires bien renseignées. Pour en juger, nous comparons ces périodes aux effectifs annuels des militaires du contingent du ministère de la Défense. Les individus des générations sous-représentées se voient éventuellement imputer une période militaire s'ils disposent d'une période d'inactivité dans leur carrière entre 18 et 30 ans.

En 2004, lors de la construction de Prisme, le choix de **projeter les carrières sur un pas trimestriel** a été fait, alors que celui des autres modèles était annuel. Ce choix, rendant la modélisation plus complexe, a été dicté par l'importance des événements infra-annuels sur le montant et sur la date de départ en retraite. Par exemple, une période d'un an de chômage n'a pas le même effet si elle couvre une année civile entière ou si elle est à cheval sur deux années. En effet, dans le premier cas, le salaire annuel moyen, paramètre important du niveau de la pension, ne sera pas affecté. Dans le deuxième cas en revanche, les salaires des deux années seront plus faibles et pourront jouer à la baisse pour le salaire annuel moyen s'ils font partie des 25 meilleurs salaires.

La dernière étape consiste donc à placer les différents reports (activité au régime général, dans un autre régime, maladie, chômage, invalidité) par trimestre à l'intérieur d'une année. Ainsi, pour chaque période, nous disposons de l'information relative au type de report ainsi que le nombre de trimestres validés associés à ce report. Ces reports sont affectés à un trimestre à l'aide de l'information « date à date »

(c'est-à-dire les dates de début et de fin d'activité reportées dans les DADS) lorsque celle-ci est disponible. Dans la mesure où, dans Prisme, chaque trimestre civil ne contient qu'un seul type de report, quatre types de reports différents peuvent être au maximum retenus par année. Cela a nécessité de mettre en place des règles afin de prioriser les différents reports suivant l'information qu'ils contiennent. Lorsque l'information « date à date » n'est pas disponible, les reports sont placés afin de minimiser les transitions en accolant les reports de même nature indépendamment du nombre d'états validés chaque année.

À l'issue de cette phase de complétion, la table initiale créée comporte autant d'observations que d'individus. Chaque individu est repéré par un numéro d'ordre et l'ensemble de ses caractéristiques est enregistré dans des variables relatives à la démographie, la carrière et la retraite, soit un total de plus de 800 variables. La partie « carrière » regroupe des variables relatives au début de carrière (âge de début d'activité, âge de fin d'études), des variables précisant le type de report et le nombre de trimestres associés pour chaque année et trimestre de la carrière, ainsi que les montants bruts des salaires annuels et les salaires forfaitaires au titre de l'AVPF.

La modélisation des carrières

Une fois la table initiale du modèle complétée des informations manquantes, les transitions de carrière doivent être estimées afin d'affecter à chaque année et à chaque trimestre projetés un statut d'activité ou d'inactivité aux individus en âge de travailler et qui ne sont pas partis en retraite.

Le module « carrière » permet de modéliser les statuts d'activité. Il permet ainsi de simuler le partage de la population entre actifs et inactifs et, au sein des actifs, entre salariés du régime général, salariés d'autres régimes (alignés et non alignés), chômeurs, salariés en arrêt maladie ou en invalidité. Ainsi, à chaque pas trimestriel de la projection, l'une des sept situations suivantes est affectée à chaque individu en âge de travailler (de 14 à 69 ans) :

- en activité au régime général (RG) ;

15. Distributions de durées tous régimes issues de l'EIC.

16. Cette cible repose sur les hypothèses de cotisants transmises par le COR pour les projections de long terme pour les années les plus récentes et sur les chiffres de la compensation pour les plus anciennes.

- en activité dans d'autres régimes alignés¹⁷ (ARA) ;
- en activité dans d'autres régimes non alignés¹⁸ (ARNA) ; au chômage ;
- en arrêt maladie avec indemnités journalières ;
- en invalidité ou accident du travail ;
- dans une autre situation ne permettant pas de valider de trimestre.

Ces états au cours de la carrière sont attribués aux individus en fonction de leurs caractéristiques individuelles et de leur trajectoire passée à l'aide d'un enchaînement d'équations logistiques dichotomiques. Le choix d'utilisation d'équations logistiques emboîtées au lieu d'une modélisation logistique multinomiale est motivé par la nécessité de contrôler à chaque pas de projection les effectifs se répartissant entre les différents types de reports et permet ainsi de faciliter les calages effectués en vue de respecter les données de cadrage fournies par le ministère chargé des Affaires sociales et de la Santé ou le COR (cf. infra, développement sur les difficultés rencontrées et les calages).

Enchaînement en projection

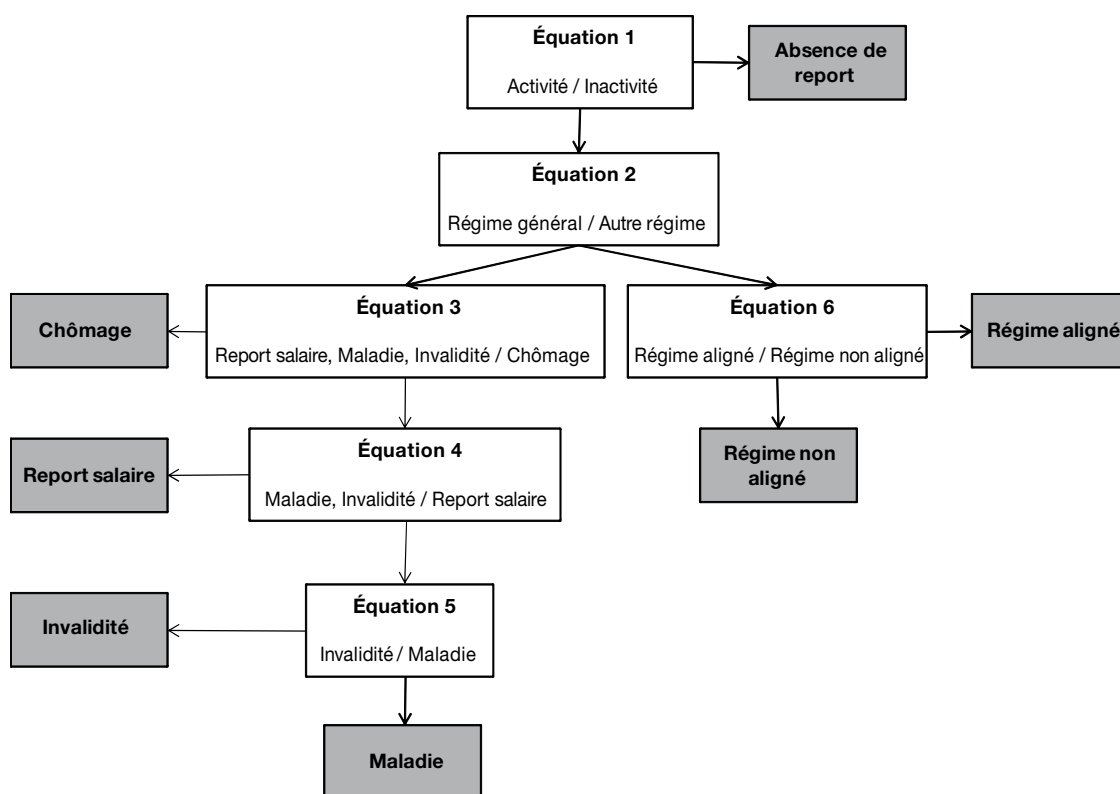
L'enchaînement des équations logistiques est schématisé par la figure II.

À chaque transition, deux situations exclusives sont testées : la probabilité d'être ou non dans tel ou tel état sur le trimestre en cours est comparée à un aléa¹⁹ qui permet soit d'imputer la situation de référence, soit d'enchaîner avec le calcul de la probabilité de transition de la situation suivante.

Par exemple, en projection, l'individu en âge de travailler rentre automatiquement dans l'enchaînement des équations présenté par la figure II et donc dans la première équation. Cette première équation calcule sa probabilité d'être en

17. Les régimes alignés sur le régime général sont la MSA salariée et le RSI (ex-AVA et ex-Organic).
 18. Les régimes non alignés sont tous les régimes hors régime général, MSA salariée et RSI. Les plus importants sont la Fonction publique d'État, la CNRACL, la MSA exploitants et les régimes des professions libérales (CnavPL). Ils comprennent également les périodes validées à l'étranger.
 19. L'aléa est tiré d'une loi uniforme entre 0 et 1.

Figure II
Enchaînement des équations logistiques



Lecture : les cases blanches correspondent aux différentes situations testées. Les cases grisées sont les reports affectés comme résultat de chaque test.

inactivité (situation de référence) contre celle d'être en activité (activité au régime général ou dans un autre régime) ou encore de valider une période assimilée de maladie, de chômage ou d'invalidité. Cette probabilité est ensuite comparée à un aléa. Cette comparaison permet d'imputer une absence de report (situation de référence) lorsque l'aléa est inférieur à la probabilité ou un report dans le cas contraire (enchaînement avec l'équation 2).

Si l'individu s'est vu attribuer de l'inactivité, il sort de l'enchaînement pour y rentrer au trimestre suivant. Dans le cas contraire, il rentre dans la deuxième équation. Cette dernière calcule sa probabilité d'être en activité dans un autre régime²⁰ (situation de référence) contre celle d'être en activité au régime général ou de valider une période assimilée²¹. Sa probabilité d'être en activité dans un autre régime est calculée puis comparée à un aléa. Cette comparaison permet d'imputer un report dans un autre régime (situation de référence) lorsque l'aléa est inférieur à la probabilité (enchaînement avec l'équation 6) ou un report au régime général ou une période assimilée dans le cas contraire (enchaînement avec l'équation 3)...

À l'issue de cet enchaînement, la situation de la personne est affectée sur le trimestre courant : elle est inactive ou assurée d'un autre régime aligné ou non aligné, ou au chômage, ou salariée du régime général, ou en maladie, ou en invalidité.

Cette opération est réalisée pour tous les individus non retraités, âgés de 14 ans à 69 ans, par genre, lieu de naissance et tranche d'âge. Elle est effectuée quel que soit de l'âge de fin d'études. Ainsi, nous considérons que les individus, au cours de leurs études, peuvent valider tous les types de reports précédemment énumérés.

En projection, lorsqu'un report d'activité au régime général est attribué à un individu dans l'année, le module « salaires » lui affecte un salaire. Le calcul des salaires futurs est essentiel : il détermine à la fois les masses de cotisations courantes et les droits à retraite futurs. Le salaire estimé est un salaire théorique annuel qui sera ensuite proratisé en fonction du nombre de trimestres au régime général l'année considérée (cf. encadré 2).

De même, la modélisation de l'assurance vieillesse des parents au foyer (AVPF) est importante pour le calcul des pensions. En effet, près de la moitié des femmes partant actuellement à

la retraite sont couvertes par ce dispositif à un moment de leur carrière. Il permet de garantir la continuité dans la constitution des droits à la retraite des personnes cessant ou réduisant leur activité professionnelle pour s'occuper de leurs enfants ou d'une personne handicapée. L'AVPF est modélisée indépendamment des transitions car elle peut être versée en complément d'une activité. L'AVPF n'est donc pas un « état » traité dans le module « carrière ». Il est considéré comme un salaire qui peut être attribué en plus d'un autre état estimé par ce module (cf. annexe).

Une fois la situation de la personne définie pour le trimestre, elle passe par le module « départs en retraite ». Ce module teste, pour chaque mois du trimestre en cours, la possibilité d'un départ en retraite. Une fois retraité, l'individu ne passe plus dans le module « carrière » et donc sa carrière n'est plus simulée (cf. encadré 3).

Estimation des équations de transitions

Les équations de transitions sont déterminées à partir des données contenues dans la table initiale du modèle de projection préalablement complétée des âges de fin d'études ainsi que des reports autres régimes (cf. supra, le développement sur la complétion et création de données relatives aux carrières). Nous utilisons un échantillon au 1/100^e en ne conservant que les individus des générations postérieures à 1934. Cette table initiale comprend, entre autres, les reports et le nombre de trimestres associés pour l'ensemble des individus. Les équations sont estimées sur les transitions de carrière entre 2005 et 2007.

Afin de déterminer des équations concernant des sous-populations homogènes et robustes, ces dernières se déclinent par genre, distance à l'âge de fin d'études et lieu de naissance (France ou étranger). Au total, 96 équations sont estimées. Les tranches d'âge varient en fonction de l'âge de fin d'études individuel pour prendre en compte l'allongement de la durée des études constaté ces dernières années ; elles ont été déterminées en étudiant la part des individus validant un report au régime général par âge (cf. tableau 1 et figure III).

Ces tranches d'âge correspondent à quatre périodes identifiées de la vie active d'un

20. Régime aligné ou non aligné.
21. Maladie, chômage, invalidité.

individu. La première se réfère à la période d'activité avant la fin des études (emplois d'étudiants), la deuxième aux premières années suivant l'obtention d'un diplôme (les six premières années de la vie active). La troisième tranche d'âge représente l'essentiel de la carrière des individus et se termine vers la cinquantaine, et la quatrième coïncide avec la décennie avant le départ en retraite (activité des seniors) et s'achève à 69 ans (dernière année projetée par le modèle).

Les variables explicatives retenues varient selon la transition considérée, les groupes d'âge, le genre, le lieu de naissance et le degré de significativité de chacune d'elles dans l'équation

étudiée. En raison du temps de calcul engendré par ce module en projection, nous avons pour l'instant limité les variables explicatives :

- le type d'état validé le trimestre précédent ;
- une variable indiquant le nombre cumulé de reports pour chacun des sept états possibles depuis le début de la carrière ;
- une variable d'interaction entre le type de report validé précédemment et le cumul des reports ARNA ;
- l'âge et une fonction quadratique de l'âge.

La variable ayant le plus de poids dans l'estimation des probabilités de valider tel ou tel type

Encadré 2

MODÉLISATION DES SALAIRES

Le module « salaire » permet d'attribuer, en projection, un salaire annuel aux individus pour lesquels au moins un report trimestriel d'activité au régime général est estimé dans l'année.

Le salaire estimé est un salaire théorique annuel qui sera ensuite proratisé, en fonction du nombre de trimestres passés au régime général l'année considérée, pour obtenir le salaire annuel.

Pour estimer les salaires, une régression MCO log-linéarisée est réalisée. Cette équation, inspirée de la modélisation de Mincer (Mincer, 1974), se décline pour quatre publics (hommes ou femmes, avant ou après l'âge de fin d'études). L'estimation des équations est réalisée à partir des salaires observés sur la période 1999 à 2008, issus de notre échantillon.

La formulation des équations est la suivante :

$$\ln(w_{jt}) = \alpha + X_{jt}\beta + u_j \quad \forall j, t \text{ (individu } j, \text{ année } t)$$

où α est la constante, X_{jt} le vecteur de variables explicatives et u_j le résidu individuel.

Les variables utilisées pour la modélisation des salaires avant l'âge de fin d'études sont les suivantes :

- l'âge courant et une fonction quadratique de l'âge ;
- une indicatrice de présence de l'individu dans l'échantillon l'année précédente¹.

Après l'âge de fin d'études, les variables explicatives de l'équation sont :

- l'âge de fin d'études ;
- l'ancienneté sur le marché du travail (issue de la différence entre l'âge courant et l'âge de fin d'études) et une fonction quadratique de l'ancienneté ;
- une indicatrice instantanée sur le chômage (présence de chômage l'année en cours) ;

- une indicatrice permanente indiquant la présence d'une période de chômage dans la carrière ;
- une indicatrice sur la présence d'une période d'invalidité dans la carrière ;
- une indicatrice de présence dans l'échantillon l'année précédente ;
- le lieu de naissance (France ou étranger) ;
- l'année de naissance de l'individu ;
- le nombre d'enfants².

L'estimation porte sur l'ensemble des salaires individuels, à temps partiel ou à temps complet qui regroupent des situations très différentes. Le pouvoir explicatif du modèle est donc plus faible que celui des estimations d'équations de gains traditionnelles qui portent sur des salaires à temps complet uniquement. Toutefois, les coefficients sont relativement proches de ceux des autres études réalisées sur données françaises (Debrand, 2004).

Afin de projeter non seulement un salaire moyen mais également de respecter la diversité des salaires autour de cette moyenne, l'écart constaté en 2008 entre le salaire observé et le salaire estimé est introduit dans l'équation. Pour les nouveaux actifs qui vont entrer en projection, la structure des résidus observés par déciles est reproduite, en fonction de l'âge, du genre et du lieu de naissance (France ou étranger). Pour les individus en activité avant la première année de projection, leur résidu est conservé et prolongé chaque année. Des calages sont ensuite effectués afin d'assurer une cohérence avec le salaire annuel moyen par tête.

1. Indicatrice destinée à prendre en compte les effets de sélection liés à l'utilisation d'un panel non cylindrique.
2. Uniquement pour les femmes.

d'état est le « type d'état validé le trimestre précédent ». Les variables indiquant le nombre cumulé de chaque report permettent d'introduire un effet de mémoire. C'est le principal

avantage de cette modélisation par rapport à l'usage d'une matrice de transitions entre états. La variable d'interaction entre l'état précédent et le cumul des reports ARNA a été introduite

Encadré 3

L'ESTIMATION DES DÉPARTS EN RETRAITE

Le module « départs en retraite » de Prisme permet de déterminer pour les individus la date du départ en retraite (cf. figure 1). Les événements simulés dans ce module reposent sur des équations logistiques estimées sur les derniers départs observés. Les informations de carrière rentrent naturellement en compte dans le départ en retraite, en particulier les données concernant les durées totales validées et cotisées tout au long de la carrière, la situation de l'assuré le trimestre précédent (en emploi, au chômage, sans validation...). Ainsi, le module « départs en retraite » estime, mois après mois, si l'assuré va ou non partir en retraite. S'il part, sa carrière n'est plus simulée et l'assuré en question ne passe plus dans le module « carrière ».

Pour les assurés ayant acquis un droit au régime général, le module « départs en retraite » est divisé en trois parties, correspondant aux trois types de départs estimés dans Prisme (départs en retraite anticipée pour carrière longue, départs à partir de l'âge légal au titre de l'inaptitude et les autres départs à partir de l'âge légal).

Départ en retraite anticipée pour carrière longue

Si l'assuré remplit les conditions nécessaires à un départ en retraite anticipée, une probabilité de départ va être estimée. Cette probabilité est calculée à l'aide de régressions logistiques déclinées par âge exact (au mois près) et genre. Parmi les variables explicatives figurent la durée en mois depuis laquelle l'assuré remplit les conditions pour un départ anticipé, sa situation vis-à-vis de l'emploi le mois précédent (activité au régime général, dans les autres régimes...), son caractère polycotisant, son pays de naissance, la présence d'une période de chômage au cours des 12 derniers mois, une indicatrice temporelle... La probabilité est comparée à un aléa tiré d'une loi uniforme entre 0 et 1, si l'aléa est inférieur à la probabilité, alors l'assuré part en retraite anticipée.

Départ à partir de l'âge légal

Lorsqu'un assuré n'est pas retraité et qu'il a atteint l'âge légal, le modèle va déterminer, chaque mois, sa probabilité de départ. Cette étape est composée de deux parties, Prisme simule dans un premier temps l'inaptitude au travail et, dans un second temps, pour les individus non inaptes, la décision de départ.

La simulation de l'inaptitude n'est effectuée que pour les assurés de moins de 65 ans. Elle repose sur un modèle logistique différencié par genre et âge. Les variables explicatives sont notamment la différence entre la durée requise pour l'obtention d'une pension à taux plein et la durée d'assurance, des indicatrices

renseignant la présence d'une période maladie et d'une période chômage au cours des 12 derniers mois, le caractère polycotisant... Les individus classés en inaptitude partent aussitôt en retraite.

Pour les individus restants, la probabilité de départ se fonde sur une équation logistique spécifique à un âge exact et un genre. Chacune des équations est estimée séparément en modifiant la liste des variables explicatives de manière à trouver la meilleure spécification possible. Une des variables déterminantes est la distance au taux plein qui mesure la différence entre la durée requise pour l'obtention d'une pension à taux plein et la durée d'assurance totale de l'assuré. Les autres variables pouvant intervenir sont : la situation vis-à-vis de l'emploi, la durée d'assurance au régime général, le caractère polycotisant, la présence d'une période de chômage et de maladie au cours des douze derniers mois, un ratio entre la durée de cotisation totale et la durée d'assurance totale, l'âge de fin d'étude, le pays de naissance, le mois en cours et une moyenne des meilleurs salaires annuels.

Application de la réforme 2010

Les départs dans Prisme sont régis par des équations logistiques estimées sur les départs observés entre 2005 et 2008. En conséquence, l'hypothèse implicite sur les comportements de départs en projection est une reproduction de cette période. Cependant, comme elle est antérieure à la réforme des retraites de 2010, des ajustements ont été nécessaires afin de tenir compte des décalages de l'âge légal et de l'âge du taux plein. Ainsi, depuis la réforme de 2010, les individus pouvant partir à 60 ans ne partent plus à l'âge légal mais en retraite anticipée.

Pour les assurés liquidant hors réforme entre 60 ans et le nouvel âge légal, la probabilité de liquider à l'âge légal correspond au « cumul » des probabilités de départ hors réforme entre 60 ans et le nouvel âge légal. Pour les départs à 58 ans (nouvel âge minimum de départ en retraite anticipée) et pour les départs à 67 ans (nouvel âge d'obtention du taux plein), on applique le même mécanisme en cumulant les probabilités de départ hors réforme entre 56 et 58 ans (respectivement 65 et 67 ans). Pour tous les autres âges de départ, les équations qui déterminent la probabilité instantanée de liquider ont été translatées, de sorte qu'une personne de 61 ans avant réforme aura la même probabilité à 63 ans après réforme, si ses caractéristiques individuelles restent proches. Néanmoins, si ces deux années supplémentaires lui permettent d'acquiescer le taux plein, l'assuré verra sa probabilité de liquider fortement augmenter sous l'effet de la variable indicatrice qui détermine si l'assuré bénéficie du taux plein ou non.

afin de tenir compte de la spécificité des carrières dans les régimes non alignés (carrières continues, peu de chômage). Cela permet d'intégrer un effet différent de la variable « cumul des reports ARNA » selon la valeur prise par le type de report précédent.

Les équations sont déclinées pour chaque tranche d'âge, genre et lieu de naissance. Cependant, en raison de problèmes de significativité pour certaines sous-catégories, des équations ont dû être regroupées afin de rendre le modèle plus robuste. Pour ces modèles, les variables « genre » et « lieu de naissance » ont été introduites²².

À titre d'exemple, le tableau 2 donne les coefficients de l'équation de transition vers le chômage pour les hommes nés en France, issus du modèle logit. Dans cette équation, la probabilité de valider un report chômage versus un report d'activité salariée ou maladie ou invalidité est calculée. Les variables non renseignées sont des variables non significatives ou des variables dont l'ajout n'améliore pas les qualités prédictives du modèle.

Pour la troisième tranche d'âge, toutes les variables sont significatives au seuil de 1 %. Les

résultats sont conformes à l'intuition. Le fait de valider tout autre type d'état le trimestre précédent (activité salariée au régime général, dans un régime aligné ou non aligné, maladie, invalidité ou inactivité) a des effets négatifs sur la probabilité d'être au chômage le trimestre donné. Par exemple, le fait d'avoir un report d'activité au régime général plutôt que du chômage le trimestre t-1 diminue la probabilité d'être au chômage le trimestre t de 99,3 % ($100 \times (0,007 - 1)$). Lorsque le cumul d'activité au régime général augmente d'un trimestre, le risque d'être au chômage baisse de 1,3 % ($100 \times (0,987 - 1)$). Le modèle apparaît correctement spécifié. En effet, les paires concordantes sont de l'ordre de 94 %.

Difficultés rencontrées et calages

Si l'application des six modèles logistiques dichotomiques reste aisée puisque chacun d'eux ne fait intervenir que deux situations exclusives,

22. Actuellement, aucune variable prenant en compte l'existence d'enfants, pour les femmes n'est introduite en raison d'un temps de calcul du modèle assez élevé. Dans un souci d'amélioration du modèle, il est envisagé de développer le lien fécondité-carrière dans le module « carrière ». Aucune variable de stratification sociale n'est également présente car cette information n'est pas disponible dans nos bases de gestion.

Tableau 1
Tranches d'âge retenues pour l'estimation des équations

Tranche d'âge 1	[14 ans ; âge de fin d'études]
Tranche d'âge 2]âge de fin d'études ; âge de fin d'études + 6 ans]
Tranche d'âge 3]âge de fin d'études + 6 ans ; âge de fin d'études + 37 ans]
Tranche d'âge 4]âge de fin d'études + 37 ans ; 69 ans]

Source : Prisme, Cnav.

Figure III
Part des individus ayant un report de salaire au régime général par âge (génération 1940)



Lecture : 40 % des individus de la génération 1940 ont un report salaire l'année de leurs 30 ans.

Champ : assurés de la génération 1940

Source : SNGC, Cnav.

l'exercice se complexifie lorsqu'il s'agit d'enchaîner ces modèles. En effet, l'adoption de ce type de modélisation est délicate car la dépendance aux conditions initiales est grande. Il est donc nécessaire de surveiller l'enchaînement des équations et les effectifs concernés à chaque étape, sous peine de dérive dans les reports présents (sureprésentation d'une catégorie voire disparition d'autres catégories...).

Prenons l'exemple de l'équation logistique définie pour déterminer le nombre de personnes inactives à chaque trimestre de la projection. Si la modélisation de cette équation n'est pas adaptée, le nombre d'inactifs générés au départ de la

projection sera soit trop fort, soit trop faible, et cet effet s'accroîtra au cours de la projection, le nombre d'inactifs générés pouvant même devenir irréaliste. Et même si les modèles logistiques qui suivent, simulant l'activité au régime général, dans un autre régime, le chômage, la maladie et l'invalidité, sont appropriés, le nombre de personnes actives passant par ces équations n'étant pas correct, les effectifs pour les transitions suivantes en seront mécaniquement affectés. En effet, si l'un des modèles dichotomiques n'est pas ajusté, les enchaînements de transitions en aval de celui-ci risquent d'engendrer de mauvaises répartitions d'effectifs. Pour contrer ces difficultés, de nombreux

Tableau 2
Coefficients estimés de la probabilité de valider un report chômage en t (équation 3) – Hommes nés en France

		Tranche d'âge			
		[14 ans ; AFE] (1)]AFE ; AFE + 6] (1)]AFE + 6 ; AFE + 37] (1)]AFE + 37 ; 69] (1)
Constante		1,208 ***	1,0985 ***	0,4346 ***	- 2,1377 ***
Âge		- 0,01869 (0,981) **	- 0,01099 (0,989) **	0,03092 (1,031) ***	0,07938 (1,083) ***
Variables de cumul	Cumul des reports activité salariée au RG	- 0,03102 (0,969) ***	- 0,03209 (0,968) ***	- 0,01355 (0,987) ***	0,0012 (1,001) **
	Cumul des reports maladie	.	.	0,00784 (1,008) ***	- 0,0206 (0,980) ***
	Cumul des reports chômage	0,14218 (1,153) ***	0,118 (1,125) ***	0,03336 (1,034) ***	0,02898 (1,029) ***
	Cumul des reports activité dans un régime aligné	.	.	- 0,0019 (0,998) ***	.
	Cumul des reports invalidité	.	.	- 0,05022 (0,951) ***	- 0,05345 (0,948) ***
	Cumul des reports activité dans un régime non aligné	.	0,01923 (1,019) ***	- 0,00251 (0,997) ***	.
Indicatrices	Malade en t-1	- 2,60733 (0,074) ***	- 2,46561 (0,085) ***	- 3,87381 (0,021) ***	- 5,46601 (0,004) ***
	Chômeur en t-1	Réf.	Réf.	Réf.	Réf.
	Activité salariée au RG en t-1	- 3,74257 (0,024) ***	- 3,74434 (0,024) ***	- 4,92837 (0,007) ***	- 6,81766 (0,001) ***
	Activité dans un régime aligné en t-1	- 4,0347 (0,018) ***	- 3,0685 (0,046) ***	- 2,7727 (0,062) ***	- 3,4662 (0,031) ***
	Activité dans un régime non aligné en t-1	- 2,75867 (0,063) ***	- 3,08416 (0,046) ***	- 2,81794 (0,060) ***	- 3,88514 (0,021) ***
	Invalide en t-1	- 3,26653 (0,038) ***	- 4,40122 (0,012) ***	- 4,91876 (0,007) ***	- 5,68689 (0,003) ***
	Inactif en t-1	- 3,27829 (0,038) ***	- 3,13587 (0,043) ***	- 3,04343 (0,048) ***	- 3,80364 (0,022) ***
Paires concordantes (en %)		83,5	86,9	93,9	97,1

Lecture : ** : significatif au seuil de 5 % ; *** : significatif au seuil de 1 % ; odds ratio entre parenthèses. Pour un homme né en France de la tranche d'âge 3, le fait d'avoir un report d'activité au régime général plutôt que du chômage le trimestre t-1 diminue la probabilité d'être au chômage le trimestre t de 99,3 % ($100 \times (0,007 - 1)$). Lorsque le cumul d'activité au régime général augmente d'un trimestre, le risque d'être au chômage baisse de 1,3 % ($100 \times (0,987 - 1)$).

Champ : hommes nés en France en âge de travailler.

Source : Prisme, Cnav.

contrôles sont effectués à chaque pas de projection afin de vérifier si les répartitions par genre et tranches d'âges ne sont pas aberrantes.

En projection, les probabilités de transitions estimées ne sont pas appliquées telles qu'estimées par le module. En effet, les équations étant estimées sur les carrières observées entre 2005 et 2007, en appliquant ces probabilités en projection, nous reproduisons les répartitions observées sur cette période sans prise en compte d'hypothèses macroéconomiques imposées. Par exemple, le taux de chômage reproduirait les tendances observées sur la période 2005 à 2007.

Des calages simples sont donc effectués afin d'augmenter ou de diminuer les probabilités individuelles données par chaque équation. Ces calages sont effectués année par année. La méthode de calage utilisée est celle employée dans Destinie 2 (cf. encadré 4). Cela permet de respecter les données de cadrage fournies par le ministère chargé des Affaires sociales et de la Santé ou par le COR.

Validation et résultats

Illustration des transitions de carrière en projection

L'attribution des reports en projection est fonction des probabilités estimées par les équations de transitions. Des calages modifient ces probabilités pour suivre un scénario macroéconomique. Les résultats présentés dans cette partie sont issus des projections de long terme

réalisées pour le COR en 2012 selon les hypothèses retenues pour le scénario B²³.

Le tableau 3 illustre les probabilités de transitions d'un report à un autre entre le 1^{er} trimestre et le 2^e trimestre de l'année 2040 (année pour laquelle les sentiers macroéconomiques sont stabilisés). Les probabilités de rester dans le même type de report sont très élevées. Cela illustre l'importance de la variable « type de report le trimestre précédent » dans la détermination d'une transition d'un trimestre à l'autre. Les probabilités de rester en activité (régime général, régime aligné, régime non aligné) ou en invalidité ou en inactivité sont plus élevées que celles de rester au chômage ou en maladie. Ces résultats sont conformes à l'intuition : les périodes de chômage et de maladie sont des périodes transitoires entre différentes périodes d'activité.

La figure IV présente la part de chaque report par tranche d'âge en 2040. La 1^{ère} tranche d'âge regroupant les étudiants est caractérisée par une forte inactivité (77 %). Leurs périodes d'activités sont essentiellement composées de petits emplois. La répartition des reports des individus venant d'entrer sur le marché du travail (2^e tranche d'âge) se distingue de celle de ceux ayant plus d'ancienneté (3^e tranche) par une part moins élevée de l'activité (notamment dans les autres régimes), liée à l'entrée sur le marché du travail de ces individus. La dernière tranche d'âge regroupe les fins de carrière qui

23. Retraites : perspectives 2020, 2040 et 2060, 11^e rapport du COR.

Encadré 4

MISE EN ŒUVRE DES ÉQUATIONS DE TRANSITIONS : IMPUTATION DE REPORTS ET CALAGES

Pour une transition donnée, notons $P_{A,X}$ et $P_{A,Y}$ les probabilités estimées par l'équation i de passer dans les états X et Y . Nous tirons un aléa issu d'une loi uniforme entre 0 et 1, que l'on compare à $P_{A,X}$.

S'il est inférieur, l'individu est dans l'état X , s'il est supérieur, il est dans l'état Y . Ces probabilités de transitions sont calées au niveau individuel afin d'obtenir les effectifs imposés pour chaque type de report année après année.

Pour faire varier les effectifs de l'état X l'année A , la probabilité $P_{A,X}$ subit la transformation :

$$P'_{A,X} = \frac{\alpha P_{A,X}}{1 + (\alpha - 1) P_{A,X}}, \quad \alpha > 0$$

Dans le cas où $\alpha > 1$, on augmente ainsi les probabilités de transition dans l'état X , et donc mécaniquement les effectifs décomptés dans cet état cette année là. À l'inverse, un coefficient $\alpha < 1$ permet de diminuer le nombre d'individus dans l'état X cette année là.

Cette formule présente l'avantage de maintenir la nouvelle probabilité calculée entre 0 et 1, mais rend légèrement plus complexe la procédure de calage, l'utilisation du coefficient α étant non linéaire.

sont généralement marquées par une part plus élevée de l'inactivité.

Validation de la carrière projetée : le cas des individus proches de la retraite

Pour vérifier la cohérence des carrières projetées avec la réalité, nous comparons les carrières des individus présents au Système National Gestions Carrières (SNGC) avec celles que le modèle Prisme leur aurait estimées. Il est délicat de comparer la carrière des assurés encore actifs

car nous ne disposons de la carrière complète des individus qu'une fois retraités. En effet, si les données concernant l'activité au régime général, les périodes maladie et chômage sont alimentées en continu, dans les fichiers de gestion de la Cnav, une partie des informations relatives aux autres régimes n'est connue qu'au moment du départ en retraite.

Dans un premier temps, nous allons comparer les carrières moyennes des individus partant en retraite en 2012. Notre analyse se concentre sur les assurés nés en 1947, 1951 et 1952, ces

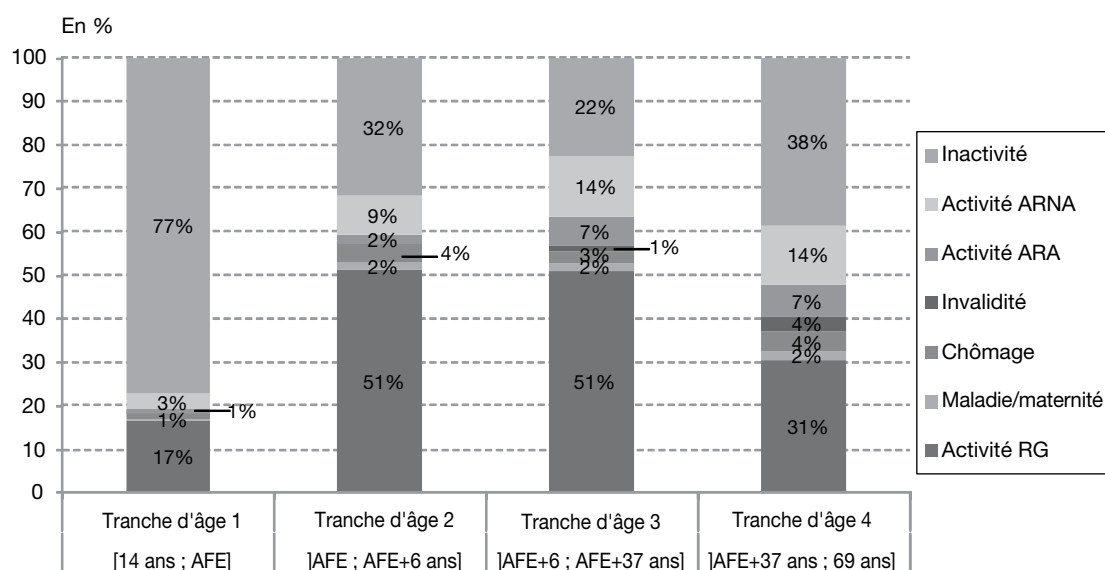
Tableau 3
Probabilités de transition entre chaque type de reports
(calculées *a posteriori* sur des données projetées)

En %

		Situation au 2 ^e trimestre 2040						
		Maladie/ maternité	Chômage	Activité RG	Activité ARA	Activité ARNA	Invalidité	Inactivité
Situation au 1 ^{er} trimestre 2040	Maladie/ maternité	43,8	6,0	34,5	0,3	0,3	4,0	11,1
	Chômage	2,6	59,0	28,4	0,7	0,4	1,1	7,9
	Activité RG	1,6	1,9	91,8	0,3	0,3	0,0	4,1
	Activité ARA	0,1	0,3	1,4	92,2	0,1	0,0	6,0
	Activité ARNA	0,0	0,1	0,7	0,0	96,4	0,0	2,7
	Invalidité	0,9	1,9	4,1	0,2	0,1	90,0	2,8
	Inactivité	0,0	0,1	1,1	0,1	0,1	0,0	98,5

Lecture : en moyenne, en 2040, 59 % des individus au chômage au 1^{er} trimestre restent au chômage au 2^e trimestre.
Champ : population en âge de travailler.
Source : Prisme, Cnav.

Figure IV
Part de chaque type de reports par tranche d'âge en projection (en 2040)



Lecture : en 2040, parmi les individus les plus jeunes (tranche d'âge 1), 77 % sont en inactivité.
Champ : Population en âge de travailler.
Source : Prisme, Cnav.

générations étant les plus représentées parmi les départs en retraite de 2012. Cette comparaison est effectuée sur les années précédant le départ en retraite (de 2007 à 2011), les projections effectuées à l'aide de Prisme débutant en 2007.

Le premier indicateur compare le **nombre moyen de trimestres** pour chaque type de report dans la période 2007 à 2011 (cf. figure V-A). Prisme reproduit assez fidèlement les différences de profils entre les générations représentées. En effet, les individus nés en 1951 ont entre 60 et 61 ans en 2012, leur âge légal de départ en retraite étant de 60 ans pour ceux nés avant le 1^{er} juillet et 60 ans et 4 mois pour ceux nés après. Il s'agit essentiellement d'individus partis à l'âge légal ou en retraite anticipée pour carrière longue (pour ceux, nés après le 1^{er} juillet 1951 et partant avant 60 ans et 4 mois). C'est également le cas pour ceux nés en 1952 dont l'âge légal est de 60 ans et 9 mois. Les individus nés en 1947 ont 65 ans en 2012, ils ont ainsi atteint l'âge dit « du taux plein » qui leur permet de partir à la retraite au taux maximum de 50 %, même s'ils n'ont pas cotisé suffisamment longtemps. Nous nous attendons donc à ce que les fins de carrières des assurés partant dès l'âge légal soient plus complètes que celle de ceux qui doivent attendre le taux plein.

En effet, alors que le profil des individus des générations 1951 et 1952 est marqué par des fins de carrière continues avec de longues périodes d'activité, leur permettant de partir à l'âge légal sans décote ou en retraite anticipée en 2012, celui des individus nés en 1947, attendant l'âge du taux plein, est caractérisé par des fins de carrière plus heurtées avec plus de périodes d'inactivité et de chômage. Ainsi, la part des reports d'inactivité est plus élevée pour la génération 1947 : en moyenne 11,5 trimestres en réalisé et 11,2 trimestres en projeté. Les générations 1951 et 1952 ont une part d'activité au régime général plus importante que la génération 1947 : 7,9 trimestres en réalisé et 7,8 trimestres en projeté pour la génération 1951, 11,4 trimestres en réalisé et 10,3 trimestres en projeté pour la génération 1952 (les fins de carrière sont souvent caractérisées par moins d'activité).

Prisme modélise de manière satisfaisante les disparités de fin de carrière entre les hommes et les femmes (cf. figures V-B et V-C). Les femmes reportent en moyenne plus de trimestres d'inactivité que les hommes. Les hommes de la génération 1947 ont reporté en 7,7 trimestres au

titre de l'inactivité en moyenne (8,2 trimestres en projection) contre 13,3 trimestres pour les femmes (12,9 trimestres en projection).

La correspondance est moins marquée pour l'activité dans les autres régimes ce qui s'explique par un manque d'information initial pour l'activité dans les régimes alignés (ARA) et dans les autres régimes non alignés (ARNA).

Un deuxième indicateur est le **nombre de transitions trimestrielles**, c'est-à-dire le nombre de changements d'état entre 2007 et 2011 (cf. figure VI-A). Il y a une bonne correspondance du nombre moyen de transitions pour les générations 1951 et 1952. Les différences sont plus marquées pour la génération 1947.

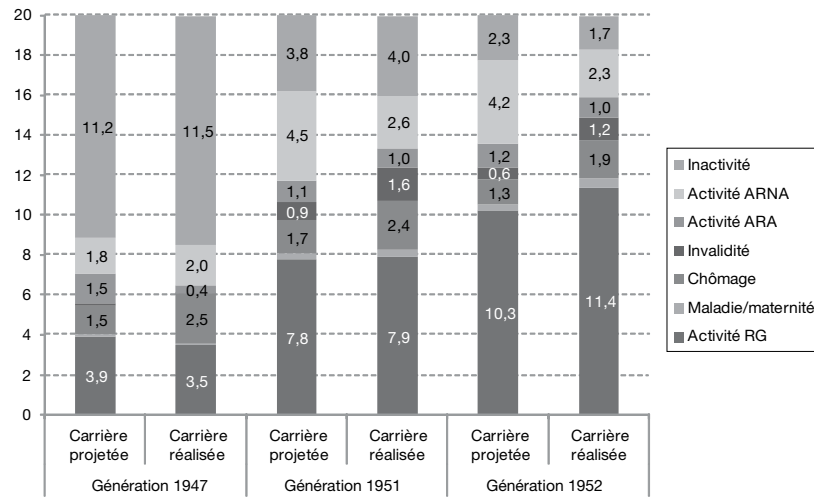
La répartition du nombre de transitions dans les carrières projetées est similaire pour les trois générations. En effet, comme cela a été présenté dans la partie modélisation des carrières, les équations de transitions sont fonction du genre, du lieu de naissance et de la tranche d'âge (déterminée en fonction de l'âge de fin d'études). Les générations 1947, 1951 et 1952 appartiennent majoritairement à la dernière tranche d'âge pendant la période étudiée. Il est donc logique que le nombre de transitions soit du même ordre de grandeur pour les mêmes tranches d'âge en projection, même si l'âge intervient en tant que variable explicative pour certaines équations de transitions.

Tandis que pour les générations 1951 et 1952, Prisme a tendance à sous estimer légèrement le nombre de transitions trimestrielles, le modèle surestimerait ces dernières pour la génération 1947 : 60 % des individus projetés ont au moins un changement d'état entre 2007 et 2011, ils sont 40 % dans la réalité. Le nombre de changements d'état trimestriel est plus faible pour les individus de la génération 1947 (individus partant au taux plein) car leur fin de carrière est souvent caractérisée par de grandes périodes d'inactivité.

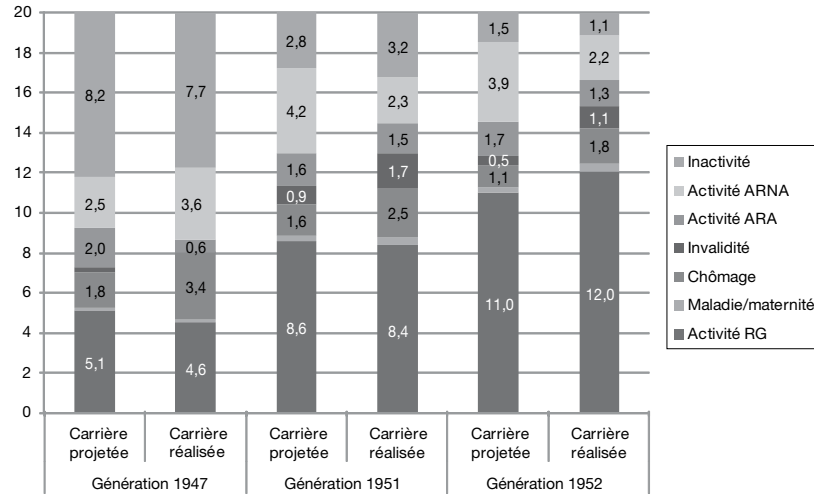
L'écart entre le nombre de changements d'état trimestriel réalisé et projeté est stable entre les hommes et les femmes (cf. figure VI-B et C).

La cohérence de la carrière projetée des assurés proches de la retraite dépend également de l'estimation des départs en retraite. Si le module « départs en retraite » n'est pas bien calibré et ne fait pas partir les « bons » assurés en retraite au « bon » moment, la comparaison entre carrière projetée et réalisée sera faussée. De plus, l'estimation des départs en retraite est elle-même

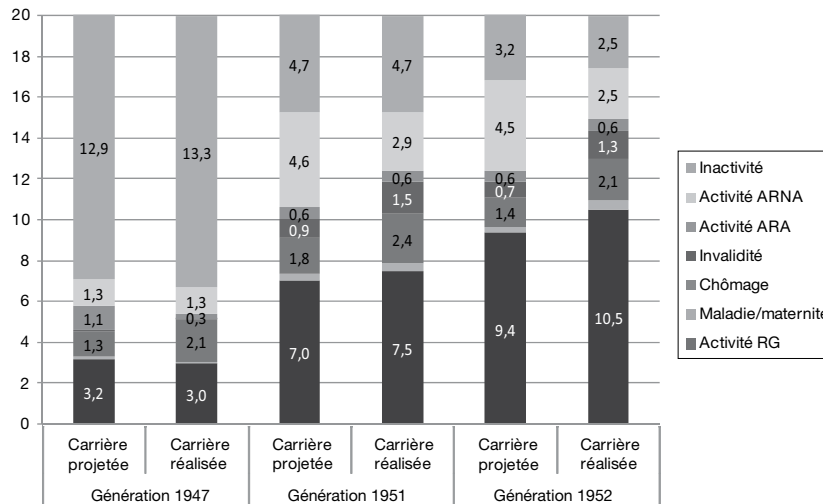
Figure V
Comparaison du nombre moyen de trimestres pour chaque report dans la période 2007 à 2011
 (générations 1947, 1951 et 1952)
A-Ensemble (hommes et femmes)



B-Hommes



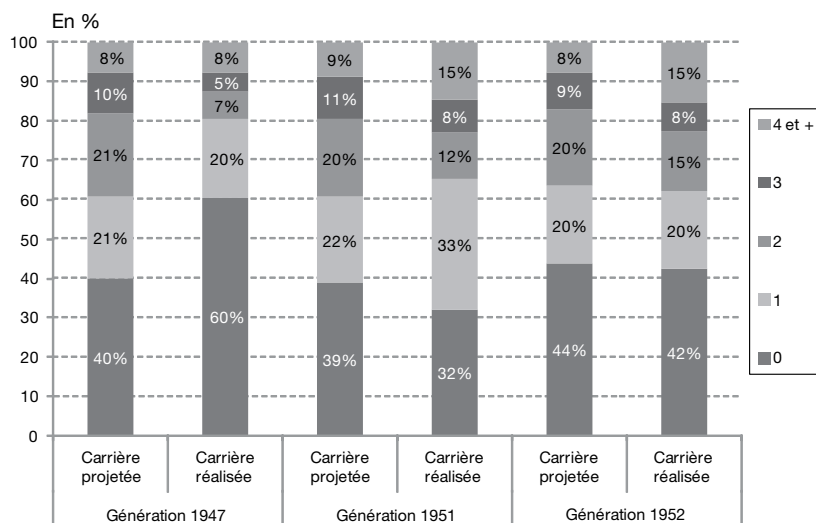
C-Femmes



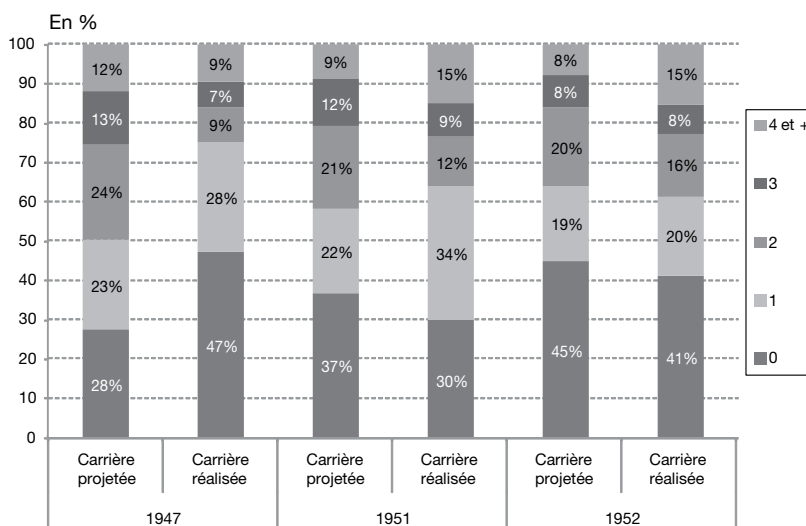
Lecture : A : pour la génération 1947, 3,5 trimestres ont été validés en moyenne au titre de l'activité au régime général. La valeur reconstituée par le modèle est de 3,9 trimestres.
 Champ : générations 1947, 1951 et 1952 partis à la retraite en 2012.
 Source : Prisme, SNGC, Cnav.

Figure VI
Comparaison du nombre de changements d'état trimestriel de 2007 à 2011
 (générations 1947, 1951 et 1952)

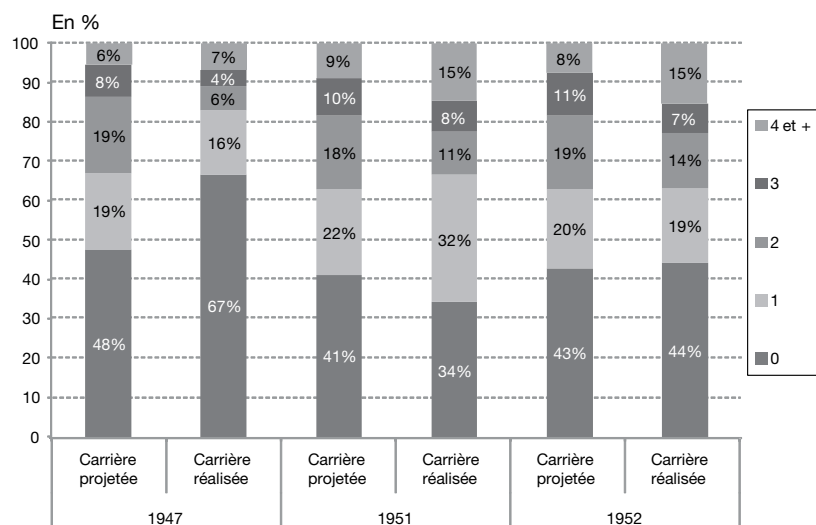
A-Hommes et femmes



B-Hommes



C-Femmes



Lecture : A : pour la génération 1952, 44 % des carrières projetées et 42 % des carrières réalisées restent dans le même état de 2007 à 2011.
 Champ : générations 1947, 1951 et 1952 partis à la retraite en 2012.
 Source : Prisme, SNGC, Cnav.

fortement liée à l'estimation des carrières, les informations concernant la carrière intervenant comme déterminant des départs en retraite (cf. encadré 2). Néanmoins, le module « départs en retraite » étant calibré pour reproduire en rétroprojection les départs constatés de 2007 à 2011²⁴, il interfère peu sur les comparaisons ci-dessus, fondées sur des indicateurs moyens.

Validation de la carrière projetée : le cas des individus en cours de carrière

Dans un deuxième temps, nous allons nous concentrer sur la carrière des générations qui ne sont pas encore retraitées en 2012. Même si l'information relative au régime général est alimentée en continu, ce n'est pas le cas de l'information relative aux autres régimes qui n'est souvent disponible qu'au moment du départ en retraite. Pour cette raison, nous nous limitons dans cette partie à la comparaison de l'activité au régime général des assurés présents au SNGC en 2012. Cette comparaison est effectuée sur les années 2007 à 2010²⁵.

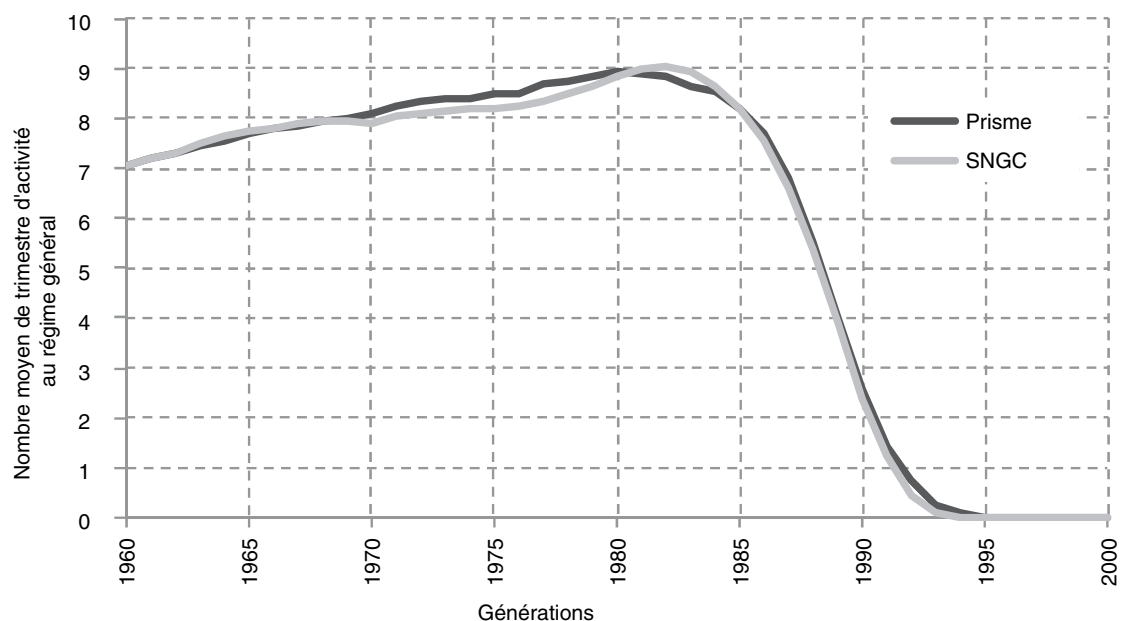
Ce troisième indicateur compare le **nombre moyen de trimestres d'activité au régime général** validés sur la période 2007 à 2010

(cf. figure VII). La comparaison des projections de Prisme avec les données disponibles au SNGC met en évidence une bonne adéquation du nombre moyen de trimestres d'activité au régime général sur la période. Les individus de la génération 1960 ont validé en moyenne 7,04 trimestres d'activité au régime général entre 2007 et 2010. Ce nombre de trimestres augmente progressivement jusqu'à la génération 1982 qui valide 9,02 trimestres en moyenne sur la période (8,87 trimestres en projection). Cette hausse s'explique par une part plus importante de l'activité au régime général en début de carrière. À partir de la génération 1983, le nombre moyen de trimestres d'activité au régime général diminue progressivement, car ces générations effectuaient leurs études ou débutaient leur activité professionnelle entre 2007 et 2010.

24. Les probabilités issues des équations de départ sont ajustées afin d'obtenir le même nombre de départs que ceux constatés au sein des différents types de départ en retraite (retraite anticipée, âge légal, inaptitude...).

25. Les reports d'activité au régime général sont totalement reportés au SNGC deux ans après leur date d'effet. Nous considérons donc que l'échantillon extrait du SNGC en 2012 comporte l'intégralité des reports d'activité au régime général jusqu'en 2010.

Figure VII
Comparaison du nombre de trimestres d'activité au régime général de 2007 à 2010



Lecture : en moyenne, les individus nés en 1985 ont validé 8,22 trimestres d'activité au régime général d'après le SNGC et 8,18 selon Prisme sur la période 2007 à 2010.
Champ : population en âge de travailler.
Source : Prisme, SNGC, Cnav.

Un dernier indicateur présente **la distribution par quartiles du nombre de trimestre d'activité au régime général**, validés entre 2007 et 2010 (cf. figure VIII). Comme précédemment, nous constatons une remarquable convergence entre les données projetées par Prisme et celles, réalisées, du SNGC. 75 % des assurés des générations 1960 à 1984 ont validé au moins 16 trimestres (soit 4 années) au régime général de 2007 à 2010. Ce nombre de trimestres diminue ensuite, comme on pouvait s'y attendre, à partir de la génération 1985.

Un exemple d'utilisation de Prisme : les projections long terme 2012

Le pôle Prévisions de la DSPR fournit régulièrement au COR des projections de long terme des retraites de la Cnav. L'exercice présenté dans ce papier est celui réalisé au deuxième semestre 2012. Plusieurs scénarios macro-économiques ont été simulés : un premier scénario, appelé scénario B et quatre autres scénarios se différenciant en termes de croissance économique.

Pour illustrer les répercussions d'un changement d'hypothèse de chômage de long terme,

nous comparons les résultats des scénarios B et C fondés sur ces hypothèses différentes :

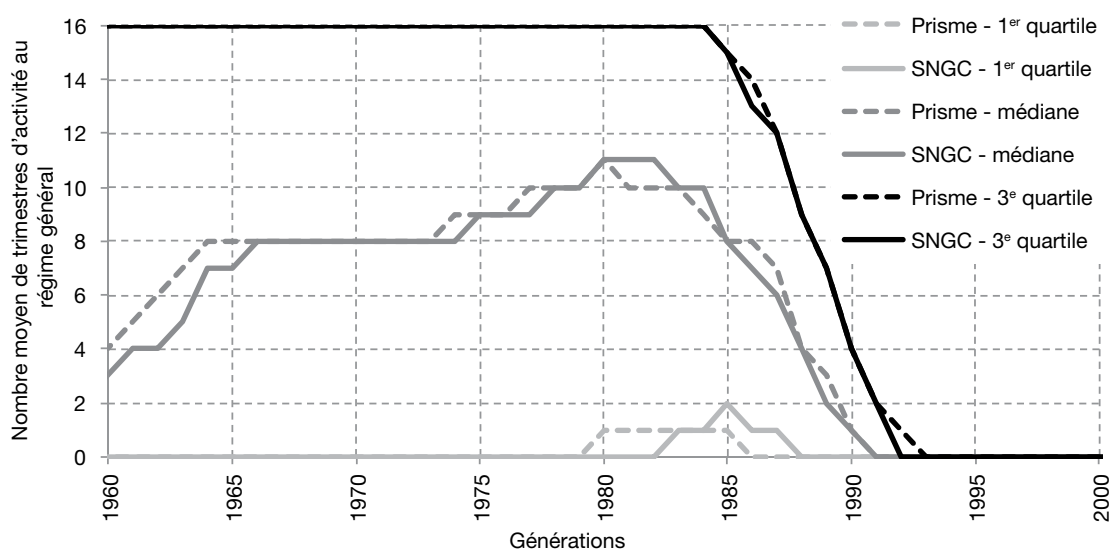
- taux de chômage de long terme : 4,5 % dans le scénario B (atteint en 2031) contre 7 % dans le scénario C (atteint en 2024) ;

- croissance annuelle de la productivité du travail à long terme : 1,5 % dans le scénario B (atteint en 2024) contre 1,3 % dans le scénario C (atteint en 2025).

Le niveau plus élevé du chômage retenu entre les deux scénarios entraînerait, très logiquement, une hausse de la part des individus ayant au moins une période de chômage dans la carrière. Cette part augmenterait de 3 points en moyenne à partir de la génération 1975 (cf. figure IX), les générations suivantes étant le plus touchées par le changement des hypothèses de chômage (elles effectuent la majeure partie de leur carrière avec un taux de chômage de 7 % au lieu des 4,5 % du scénario B).

Ce changement d'hypothèse influencerait également sur le poids du chômage dans la carrière. La part des reports de chômage augmenterait d'un point à partir de la génération 1975 (cf. figure X).

Figure VIII
Comparaison de la distribution du nombre de trimestres d'activité au régime général de 2007 à 2010 (1^{er} quartile, médiane, 3^e quartile)

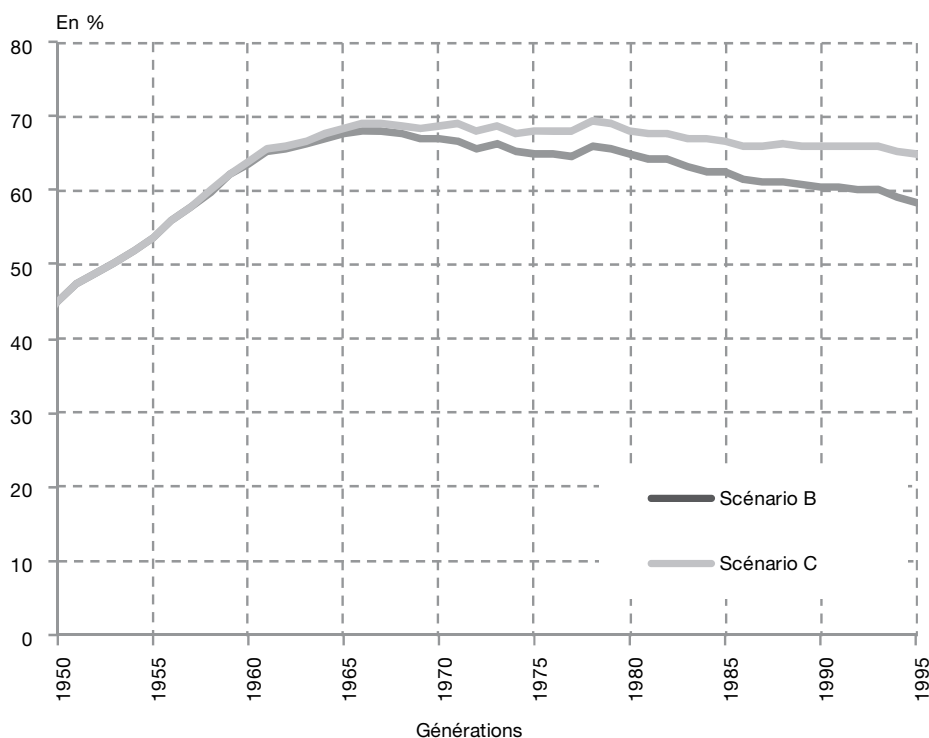


Lecture : 50 % des assurés de la génération 1980 ont validé au moins 11 trimestres d'activité au régime général sur la période 2007 à 2010.

Champ : population en âge de travailler.

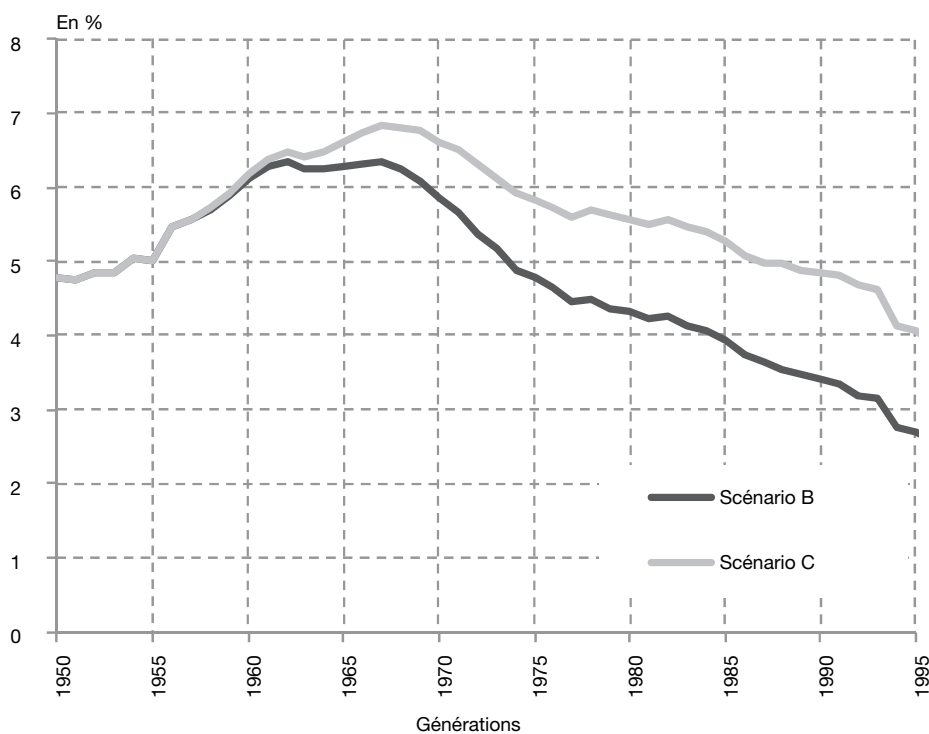
Source : Prisme, SNGC, Cnav.

Figure IX
Part des retraités ayant au moins un report chômage dans la carrière, par génération



Lecture : au moment de leur retraite, 60 % des individus nés en 1990 auraient au moins un report chômage dans leur carrière d'après le scénario B, 65 % d'après le scénario C.
Champ : projection de retraités Cnav.
Source : Prisme, Cnav.

Figure X
Part du chômage dans la carrière (pour les retraités en projection)



Lecture : au moment de leur retraite, la part du chômage dans la carrière des individus nés en 1990 serait de 3,4 % d'après le scénario B et 4,8 % d'après le scénario C.
La part du chômage est calculée en rapportant le nombre de reports chômage au total des reports entrant dans le calcul de la retraite, les trimestres d'inactivité ne sont pas comptabilisés ici.
Champ : projection de retraités Cnav.
Source : Prisme, Cnav.

Le tableau 4 reprend l'indicateur présenté par la figure IX en le détaillant par tranche d'âge pour la génération 1990²⁶. La hausse du chômage de long terme de 4,5 % dans le scénario B à 7 % dans le scénario C entraînerait une hausse de la part des individus ayant au moins une période de chômage aux âges les plus élevés (tranches d'âge 3 et 4). La tranche d'âge 4 serait la plus touchée (29,4 % contre 25,1 %). Ces résultats sont conformes à l'intuition : les fins de carrière sont fortement marquées par les hausses de taux de chômage. Les deux premières tranches d'âge sont beaucoup moins sensibles à la hausse du chômage, car l'écart du chômage entre les deux scénarios intervient surtout lorsque la génération 1990 atteint la troisième tranche d'âge²⁷. C'est durant cette période que les salariés effectuent la plus grande partie de leur carrière, ce qui explique la part élevée des individus ayant au moins un report chômage.

La figure XI compare les durées moyennes en projection de chaque période de chômage par génération. Dans le scénario C, la part des individus au chômage et la fréquence de ces durées sont plus importantes que dans le scénario B. L'écart moyen de durée en résultant est d'un mois et demi.

Les hypothèses retenues entre les deux projections en termes de baisse du nombre de cotisants et de hausse du taux de chômage n'ont que peu d'effet sur le nombre de retraités ou sur les âges moyens et durées moyennes au moment du départ. Les moindres validations de durées sous forme de salaire (baisse du nombre de cotisants) seraient, en effet, compensées par des validations supplémentaires liées aux périodes assimilées de chômage.

En revanche, les pensions moyennes annuelles des nouveaux retraités sont plus faibles dans le scénario C que dans le scénario B, conséquence

des niveaux de salaires plus faibles (hypothèse d'évolution des salaires de 1,3 % par an à terme contre 1,5 % par an et augmentation du taux de chômage), l'écart augmentant progressivement jusqu'à atteindre 820 euros en 2060 sur la pension annuelle moyenne (cf. figure XII).

Ces baisses de pension cumulées à un nombre de retraités quasiment identique se traduisent par une baisse des masses annuelles de prestations versées de l'ordre de 7,5 milliards d'euros en 2060 (3,1 % par rapport au scénario B).

* *
*

Depuis sa création, en 2004, Prisme est devenu un outil indispensable et une référence pour la simulation et l'évaluation des réformes du régime général demandées par les organismes concernés par la retraite en France (ministères, conseils, commissions...). Il est utilisé à la fois en tant qu'outil de projection de long terme et modèle de prévisions financières de très court terme. Les résultats et analyses produits constituent un élément réel d'aide à la décision pour les pouvoirs publics.

Les bases de données sur lesquelles repose le modèle Prisme sont issues de fichiers de gestion très complexes et leur utilisation à des fins statistiques requiert un important travail de préparation en amont. Grâce à la richesse de ces fichiers, une modélisation très fine des

26. Le taux de chômage est similaire entre les deux scénarios jusqu'en 2019. La génération 1990 est donc touchée par un taux de chômage supérieur l'année de ses 30 ans et effectue par conséquent l'essentiel de sa carrière, dans le scénario C, sous une hypothèse de chômage supérieure à celle du scénario B. C'est l'une des premières générations concernée par ce changement d'hypothèse.

27. L'âge de fin d'étude est en moyenne de 20,6 ans pour la génération 1990.

Tableau 4
Part des individus ayant au moins un report chômage dans la carrière par tranche d'âge pour les individus nés en 1990

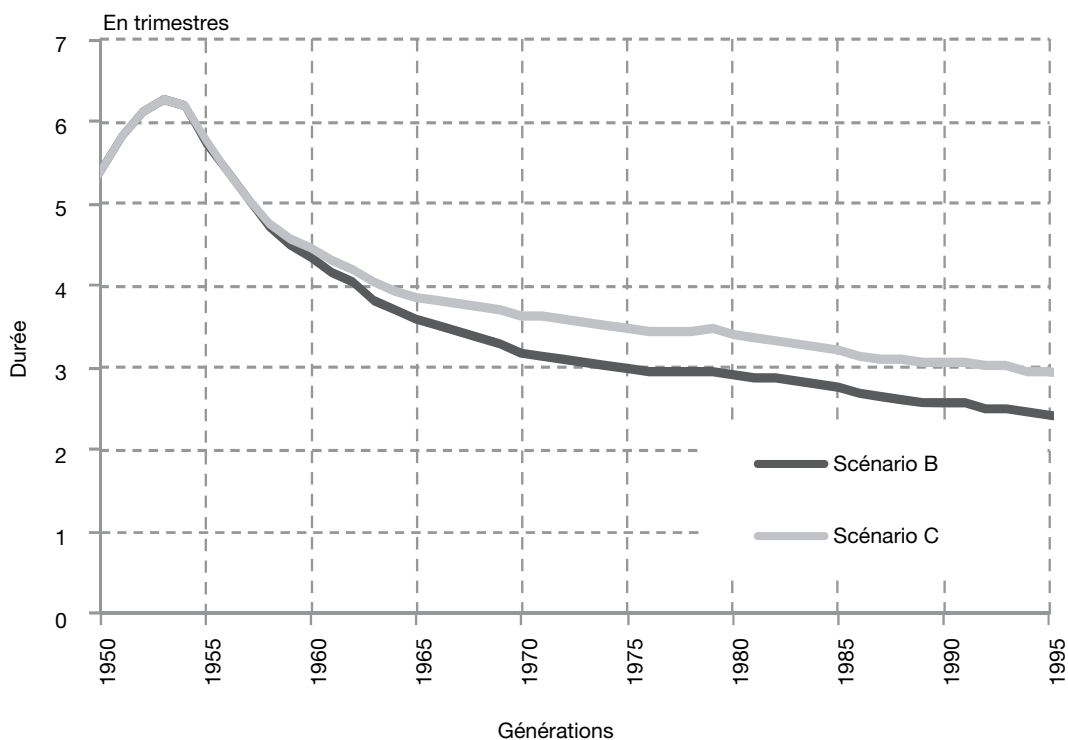
	Scénario B	Scénario C
Tranche d'âge 1 [14 ans ; AFE]	16,8	16,8
Tranche d'âge 2 [AFE ; AFE + 6 ans]	40,8	40,8
Tranche d'âge 3 [AFE + 6 ans ; AFE + 37 ans]	70,2	72,8
Tranche d'âge 4 [AFE + 37 ans ; 69 ans]	25,1	29,4

Lecture : 70,2% des individus de la tranche d'âge 3 valideraient au moins un report chômage dans le scénario B.

Champ : population en âge de travailler, projection

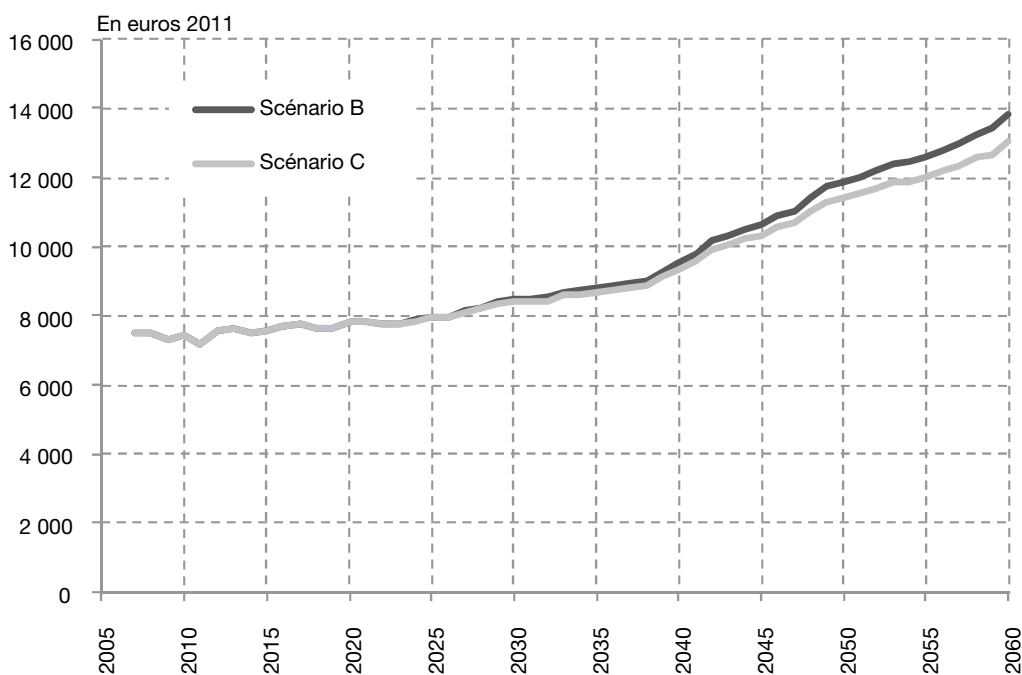
Source : Prisme, Cnav.

Figure XI
Durée de chaque période de chômage par génération



Lecture : la durée moyenne d'une période de chômage serait de 3,2 trimestres pour la génération 1970 dans le scénario B, elle passerait à 3,6 trimestres dans le scénario C.
Champ : projection de retraités Cnav.
Source : Prisme, Cnav.

Figure XII
Pension moyenne annuelle des nouveaux retraités de droit direct (en euros 2011)



Lecture : les assurés partis en retraite en 2050 perçoivent une pension moyenne annuelle de droit propre de 12000€ (€2011) dans le scénario B.
Champ : projection de retraités Cnav.
Source : Prisme, Cnav.

trajectoires professionnelles (échantillon de plus de cinq millions d'individus, pas trimestriel de la projection, etc.) est effectuée et mise à jour fréquemment pour une meilleure précision des travaux réalisés.

Prisme est un modèle en perpétuelle évolution. En effet, depuis sa création, de nombreuses améliorations y ont été apportées (remplacement de la matrice de transitions de carrière par un enchaînement d'équations logistiques, pas

mensuel des départs à la retraite, intégration du compte pénibilité, etc.). À l'avenir, Prisme a pour ambition d'affiner les carrières projetées par l'introduction d'un lien entre la fécondité et l'activité professionnelle. De plus, la création de nouveaux fichiers de gestion (répertoire de gestion des carrières unique, échanges interrégimes de retraite) devrait lui permettre d'améliorer ses estimations sur le champ du régime général mais également sur les carrières des autres régimes. □

BIBLIOGRAPHIE

Albert C., Berteau-Rapin C. et Di Porto A. (2009), « Prisme, le modèle de microsimulation dynamique de la Cnav, régime français de pension de retraite », www.statistiques-recherches.cnav.fr.

Bachelet M., Leduc A. et Marino A. (2014), « Les biographies du modèle Destinie II : rebasage et projection », *Document de travail* de la Direction des études et synthèses économiques de l'Insee, G2014/01.

Blanchet D., Buffeteau S., Crenner E. et Le Minez S. (2010), « Le modèle de microsimulation Destinie 2 : principales caractéristiques et premiers résultats », *Économie et statistique*, n° 441-442, pp. 101-121.

Chaput H. et El Mekkaoui de Freitas N. (2005), « Les durées d'assurance validées par les actifs pour leur retraite : une première évaluation à partir de l'Échantillon interrégimes de cotisants », *Études et résultats*, Drees, n° 400.

Commission des comptes de la Sécurité Sociale (2012), *Les comptes de la sécurité sociale : résultats 2011, prévisions 2012 et 2013*, Rapport d'octobre 2012, 431 pages.

Conseil d'orientation des retraites (2012), *Retraites : perspectives 2020, 2040 et 2060*, 11^e rapport du COR, 145 pages.

Debrand T. et Privat A.G. (2004), « Salaires individuels et évolutions macroéconomiques de 1947 à 2000 », *Revue de l'OFCE*, n° 89.

Drees (2005), *Échantillon interrégimes de cotisants 2005*, guide d'exploitation.

Duc C., Lequin L., Housset F. et Plouhinec C. (2013), « Le modèle de microsimulation Trajectoire, Trajectoire de carrières tous régimes », *Document de travail* de la Direction de la recherche, des études, de l'évaluation et des statistiques (Drees), n°40, mai 2013.

Insee (1999), « Le modèle de microsimulation dynamique Destinie », *Document de travail* de la Direction des études et synthèses économiques de l'Insee, G 9913.

Li J., O'Donoghue C. (2012), « A methodological survey of dynamic microsimulation models », *UNU-MERIT Working Paper Serie*, n° 2012-002.

Mincer J. (1974), *Schooling, Experience and Earnings*, Columbia University Press, New York.

Poubelle V., Albert C., Beurnier P., Couhin J., Grave N. (2006), « Prisme, le modèle de la Cnav », *Retraite et société*, n° 48, pp. 202-215.

MODÉLISATION DE L'AVPF

Créée en 1972, l'AVPF (assurance vieillesse des parents au foyer) a pour objectif de limiter les effets des diminutions ou des arrêts d'activité liés à la charge d'enfants sur les futures retraites des parents. Elle concerne les bénéficiaires de certaines prestations familiales versées par les CAF (le complément familial, l'allocation adulte handicapé, l'allocation d'éducation de l'enfant handicapé...). Elle consiste en la prise en charge de cotisations d'assurance vieillesse par la branche famille, versées à la Cnav et reportées directement sur le compte individuel de l'assuré sous la forme de salaires forfaitaires annuels AVPF et de validations de trimestres d'assurance vieillesse¹.

La perception d'un salaire AVPF est modélisée annuellement dans le modèle, indépendamment des transitions. Il peut bénéficier aux femmes comme aux hommes. Cependant, dans les faits les hommes sont peu concernés par cette allocation. L'événement AVPF n'est donc pas, dans le modèle, modélisé pour les hommes (pour ces derniers, nous ne disposons d'ailleurs pas d'informations suffisantes sur les naissances d'enfants).

La modélisation est réalisée au sein des femmes non retraitées, nées entre 1952 et 1990, ayant des enfants, qu'elles soient ou non passées par le régime général entre 1999 et 2004. Deux équations logistiques sont

1. Le nombre de trimestres validés est fonction du montant du salaire forfaitaire AVPF. Un trimestre est validé dès que le salaire forfaitaire AVPF est égal à 150 heures du Smic en vigueur au 1^{er} janvier de l'année considérée.

estimées selon que la bénéficiaire de l'AVPF ait plus ou moins de 35 ans. Les variables explicatives sont les suivantes :

- l'âge de la femme et une fonction quadratique de l'âge ;
- le lieu de naissance (France ou étranger) ;
- l'âge de fin d'études ;
- la présence d'un report d'activité l'année précédente (régime général et autres régimes) ;
- la présence d'un report chômage l'année précédente (uniquement pour les femmes de moins de 35 ans) ;
- le nombre d'enfants ;
- la présence d'un enfant de moins de 3 ans ;
- l'âge du benjamin ;
- le salaire de l'année précédente ;
- le nombre de trimestres au régime général sur la carrière ;
- la présence d'un salaire AVPF l'année précédente.

En projection, selon les caractéristiques individuelles de chaque femme, la probabilité d'avoir de l'AVPF est ainsi calculée et comparée à un aléa. Si la probabilité est supérieure à l'aléa, un salaire AVPF annuel est estimé. Ce salaire est affecté par tirage aléatoire et peut prendre trois valeurs possibles correspondant à la compensation de l'arrêt complet d'activité (salaire forfaitaire affecté permettant de valider 10 trimestres), à la réduction de l'activité (2 trimestres) ou à la réduction forte d'activité (5 trimestres).
