

CONSEIL D'ORIENTATION DES RETRAITES
Séance plénière du 5 mars 2020 à 9h30
« Point sur les modèles de microsimulation »

Document N° 12
<i>Document de travail, n'engage pas le Conseil</i>

Description du modèle de microsimulation de l'Agirc-Arrco

Direction technique, AGIRC-ARRCO

DT 2020-08
10 février 2020

Description du modèle de microsimulation de l'Agirc-Arrco

(séance COR février 2020)

Cette note présente le modèle de microsimulation de l'Agirc-Arrco. Elle est préparée en vue de la séance du COR de février 2020, consacrée au sujet de la microsimulation et aux modèles de microsimulation sur les retraites en France et à l'étranger.

La première partie introduit brièvement le régime Agirc-Arrco, ainsi que le principe et l'intérêt général de l'approche par microsimulation. La deuxième partie décrit le modèle de microsimulation de l'Agirc-Arrco.

Table des matières

INTRODUCTION	2
MODELE DE MICROSIMULATION DE L'AGIRC-ARRCO	3
1. STRUCTURE GENERALE ET CARACTERISTIQUES	3
2. ECHANTILLON/BASE INITIALE.....	4
<i>a. Les données</i>	4
<i>b. Taille et tirage de l'échantillon</i>	6
3. ARCHITECTURE DU MODELES ET FONCTIONNEMENT DES MODULES.....	7
<i>a. Architecture du modèle</i>	7
<i>b. Modules</i>	7
4. CALAGES	14

Introduction

L'Agirc-Arrco est le régime de la retraite complémentaire des salariés du secteur privé. Il concerne la quasi-totalité des salariés de l'industrie, du commerce, et des services et de l'agriculture. Il résulte de la fusion de deux régimes de retraite complémentaire, l'Agirc et l'Arrco, au 1^{er} janvier 2019.

Ce sont les partenaires sociaux, représentants des organisations patronales et syndicales qui ont créé les deux régimes de retraite complémentaire à l'origine du régime actuel.

L'Agirc¹ est le régime de retraite complémentaire des cadres. Il a été créé afin de leur permettre d'acquérir des droits sur la part du salaire au-dessus du PSS, non couverte par le régime de base.

L'Arrco² est le régime de retraite complémentaire de tous les salariés du secteur privé (cadres et non cadres). L'objectif étant que les non cadres aient une retraite complémentaire comme les cadres sur la totalité de leur salaire (dans la limite de trois PSS). Le régime Arrco porte également sur la partie inférieure au PSS pour apporter une retraite complémentaire aux cadres et non-cadres.

A compter du 1^{er} janvier 2019, ces deux régimes de retraite complémentaire ont fusionné en un seul régime, le régime unifié Agirc-Arrco.

C'est un régime paritaire piloté par les organisations patronales et syndicales. Il fonctionne par répartition et par points. Les cotisations des actifs (à la charge de l'employeur et du salarié) ouvrent des droits de retraite sous forme de points, qui sont liquidés au moment de départ à la retraite. Ces cotisations servent également chaque année au financement des pensions des retraités, ainsi qu'à la constitution des réserves du régime permettant de faire face aux évolutions démographiques et aux aléas économiques.

Le modèle de microsimulation de la retraite de l'Agirc-Arrco présenté dans cette note est en cours d'élaboration et ne produit pas encore des résultats. Ce modèle servira à projeter les dépenses d'allocations du régime Agirc-Arrco, ainsi que ses ressources, et sera employé pour des projections de court, moyen et long terme. Les projections vont répondre aux besoins de pilotage du régime et les possibilités induites par la microsimulation permettront de répondre aux demandes des administrateurs du régime, ainsi que des institutions comme le Conseil d'orientation des retraites ou la Direction de la sécurité sociale. Le modèle sera également utilisé pour effectuer des prévisions de volume d'activité apportant une aide au pilotage des charges de gestion.

Approche de la microsimulation : intérêt et principe général

Le principe de la microsimulation est de permettre de tenir compte des hétérogénéités au sein de la population en s'intéressant aux comportements des agents individuels. Ce principe s'oppose aux approches agrégées, qui sont basées sur l'hypothèse d'un agent représentatif d'une population donnée.

¹ La Convention collective nationale de retraite et de prévoyance des cadres a institué le régime Agirc le 14 mars 1947.

² L'Accord national interprofessionnel de retraite complémentaire a institué le régime Arrco le 8 décembre 1961.

La microsimulation offre ainsi la capacité de simuler des phénomènes et des interactions complexes, que l'approche macroéconomique ne peut pas modéliser, et dans lesquels les comportements et les caractéristiques individuels influencent clairement les résultats.

Dans un contexte où la législation et la réglementation du système de retraite français sont complexes, avec les spécificités des différents régimes, et où des réformes des retraites successives ont eu lieu, le recours à ce principe de projection par microsimulation, notamment dynamique, est nécessaire. Il y a un besoin pour que les principaux régimes de retraite modélisent leurs trajectoires professionnelles et les départs en retraite d'une façon plus fine. En ce qui concerne l'Agirc-Arrco, les coefficients temporaires de solidarité et de majoration ou le plafonnement des majorations familiales sont des exemples de réglementations particulières au régime, dont l'incidence sur le montant des pensions et le départ en retraite n'est correctement évaluée qu'au niveau individuel. Le dispositif de retraite anticipée pour carrières longues (aux règles relativement complexes) est un autre exemple illustrant l'apport d'un modèle de microsimulation. Cela aide à contribuer d'une manière plus précise au pilotage du système et à une meilleure évaluation des réformes des retraites - que ce soit pour le régime Agirc-Arrco particulièrement, ou pour le système de retraite en général.

Les limites des modèles actuels de projection (le modèle Agirc pour les cadres et le modèle Arrco pour l'ensemble des salariés du secteur privé) qui sont des modèles retenant une approche agrégée, ont conduit aujourd'hui au développement d'un modèle de microsimulation. Ce projet a été permis par la disponibilité des données indispensables à la microsimulation et dont l'Agirc-Arrco ne disposait pas auparavant, ainsi que par un développement des moyens informatiques.

Modèle de microsimulation de l'Agirc-Arrco

1. Structure générale et caractéristiques

Le modèle de microsimulation de la retraite de l'Agirc-Arrco est un modèle de microsimulation dynamique, dans la mesure où les caractéristiques des individus sont simulées au fil du temps et où leurs comportements évoluent à court, moyen et long terme.

Au point de départ de la projection (la période initiale), un stock d'individus est disponible et leurs caractéristiques sont connues à cette période initiale. Ensuite, le modèle simule les trajectoires et les comportements de ces individus (notamment la carrière et le départ en retraite) et l'évolution de leurs caractéristiques. Les caractéristiques des individus ne sont donc pas figées et peuvent éventuellement être modifiées à chaque période. Le modèle simule également les entrées de nouveaux individus et les sorties définitives d'autres.

Le recours à ce type de modèle dynamique est indispensable pour la simulation des retraites, où les comportements individuels au cours du temps jouent un rôle important et influencent les résultats du modèle. Le départ à la retraite et le montant des pensions reçues ne sont pas le résultat des situations individuelles courantes, mais elles dépendent de l'intégralité de l'historique professionnelle des individus.

Par exemple, l'évolution de la situation professionnelle en fin de carrière peut avoir une influence sur le comportement de départ que l'on ne capterait pas si les individus restaient dans le même état tout au long de leur carrière. C'est un modèle dit de période, qui suit une boucle par année - individu, dans lequel la base initiale est actualisée pour l'ensemble des individus à chaque année de la projection en général, avec une précision mensuelle l'année de départ en retraite.

C'est un modèle « régime », dans le sens où la population de « référence » modélisée est identifiée par son affiliation à l'Agirc-Arrco : il s'agit ainsi de l'ensemble des individus cotisants ou ayant cotisé au moins une fois à l'Agirc-Arrco au cours de leur carrière et/ou bénéficiant d'un droit dérivé Agirc-Arrco (i.e. cotisants, radiés, allocataires de droit direct, allocataires de droit dérivé).

Il s'agit également d'un modèle « ouvert » (par opposition aux modèles « fermés »). Le modèle simule une population de « référence » et une population « périphérique », qui sert uniquement à décrire les ayant-droits³ de cette population de référence : les conjoints et les enfants ne font pas l'objet d'une simulation au-delà des droits dérivés qu'ils ouvrent et des majorations familiales attribuées pour enfants.

2. Echantillon/Base initiale

a. Les données

Certaines données sont directement connues via le système d'information de l'Agirc-Arrco, tandis que d'autres nécessitent une complétion, soit parce que ces données ne sont pas disponibles dans le système d'information de l'Agirc-Arrco, soit parce qu'elles sont présentes mais incomplètes.

Sources principales :

Les données initiales utilisées pour les cotisants sont des données en historiques, provenant de deux sources principales : les bases carrières et droits des actifs (système d'information de l'Agirc-Arrco), et les bases durées d'assurances des actifs, qui sont des données inter-régimes, issues du droit à l'information (DAI). Bien que l'Agirc-Arrco soit un régime par points et que le modèle ne simule que les affiliés du régime, les données des durées d'assurances inter-régimes sont indispensables pour la simulation des départs en retraite. Les conditions de départ en retraite sont déterminées par la législation du régime général, qui dépend des droits acquis tous régimes, d'où la nécessité des durées d'assurances inter-régimes.

Pour les allocataires, ce sont les bases carrières et droits des retraités et les bases allocataires, qui proviennent du système d'information de l'Agirc-Arrco, ainsi que les données inter-régimes de durées d'assurances des retraités (RCIV).

Les données historiques de carrières et des durées d'assurances des retraités ne sont pas utilisées en projections. Ils sont nécessaires pour des besoins de modélisation en amont.

Ces données doivent faire l'objet d'une complétion pour les besoins d'estimation et de projection des équations de transitions sur le marché du travail et de salaires.

³ Pour la simulation de la réversion notamment.

1. L'âge de fin d'études⁴ est une information importante qui détermine les trajectoires professionnelles et salariales, mais elle est indisponible dans les données de l'Agirc-Arrco.

La méthode utilisée est d'imputer l'âge de fin d'études comme étant l'année qui précède celle au cours de laquelle l'individu atteint ou dépasse, pour la première fois, le seuil de salaire équivalent à une validation de trois trimestres au régime général (reporté à un âge entre 15 et 30 ans)⁵.

Dans le cas où les individus n'atteignent pas ce seuil entre 15-30 ans ou que leur affiliation à l'Agirc-Arrco est après leur 30 ans⁶, un âge de fin d'études leur est attribué par un tirage aléatoire dans une distribution empirique, selon le sexe, la naissance en France ou à l'étranger et l'âge d'entrée sur le marché du travail. Cette distribution est obtenue à la suite de la première imputation, et elle est donc basée sur les affiliés de l'Agirc-Arrco pour lesquelles un âge de fin d'études a pu être calculé.

2. Trajectoire professionnelle

Les durées d'assurances (cotisées/validées) tous régimes, notamment les dernières années de carrière qui sont inconnues pour certaines générations (i.e. données non arrêtées aux mêmes dates selon les campagnes DAI⁷), ainsi que des données en début de carrière pour les jeunes générations, sont également complétées. Les campagnes DAI ont lieu tous les cinq ans à partir de l'âge de 35 ans et jusqu'à 65 ans. Ainsi, la dernière information sur les durées d'assurances n'est pas disponible à la même année pour toutes les générations. La dernière campagne est celle de l'année 2017 et la date initiale de projection du modèle est 2016. Les durées d'assurances pour les générations qui ne sont pas couvertes par la campagne 2017 ne sont pas disponibles entre la date de leur dernière campagne et l'année 2016. Pour les jeunes générations qui n'ont pas encore atteint l'âge de 35 ans, l'information sur leurs durées d'assurances n'est pas connue non plus.

La complétion des durées d'assurances est basée sur deux sources: la base EIC⁸ 2013 et les données issues du DAI. Pour les générations présentes dans l'EIC et dont les durées d'assurances ne sont pas complètes dans les données DAI, les données de l'EIC sont exploitées. Pour les générations ne rentrant pas dans le champ de l'EIC, les données utilisées sont celles des générations les plus proches présentes dans l'EIC ou dans la base DAI quand elles sont complètement disponibles. L'approche de complétion adoptée est de reproduire les distributions du nombre des trimestres et de leurs types observées dans ces deux sources de données, selon le sexe, l'âge, et la naissance en France ou à l'étranger.

Autres sources :

⁴ Il convient plus précisément d'appeler cet âge, l'âge d'insertion sur le marché du travail.

⁵ Le critère utilisé pour la validation d'un trimestre au régime général est depuis 1972, 200 heures Smic. La réforme des retraites de 2014 a abaissé ce seuil à 150 heures Smic.

⁶ Dans le premier cas, les individus sont affiliés à l'Agirc-Arrco entre 15-30 ans, mais n'atteignent pas le seuil dans cette tranche d'âge. Dans le deuxième cas, l'affiliation est après l'âge de 30 ans. Ce qui se passe avant n'est pas observé, il se peut que ce seuil soit atteint avant leur affiliation au régime.

⁷ Droit à l'Information.

⁸ L'Echantillon Inter-régime des Cotisants, une base de données constituée par la DREES qui couvrent une génération sur quatre entre la génération 1942 et 1954, puis une génération sur deux jusqu'à la génération 1990.

Le système d'information de l'Agirc-Arrco ne dispose pas des données sur le nombre d'enfants et le statut marital. Ces données sont estimées à partir d'une étude réalisée sur l'EDP en appariement avec le panel DADS.

Au final, l'ensemble de ces bases permet de connaître et d'utiliser les données suivantes :

- Des données démographiques : la date de naissance, le sexe, naissance en France ou à l'étranger, l'année d'affiliation, l'âge de fin d'étude, le nombre d'enfants. Concernant le statut marital, il n'est attribué qu'au moment du décès de l'individu.
- Des données de carrière : le salaire, les points acquis à l'Agirc-Arrco (selon le type de points et par tranches), les durées d'assurances (selon le type de trimestres).
- Des données de retraite : la date du début des droits, le type de droit (droits directs ou dérivés), les pensions Agirc et Arrco.

b. Taille et tirage de l'échantillon

L'échantillon initial est composé des générations nées entre 1896 et 2005 (individus âgés de 11 à 120 ans à la période initiale de projection 2016)⁹.

A ce stade du développement du modèle, la taille de l'échantillon est de 1/25^{ème} (4%) de la population totale de l'Agirc-Arrco, soit environ 2,2 millions d'individus. Cet échantillonnage est effectué selon un tirage aléatoire.

Le choix de la taille est principalement déterminé par une nécessité de précision importante, notamment pour les besoins de projections financières de court-terme, ainsi que pour l'intérêt accordé à certaines catégories de sous-populations de taille restreinte.

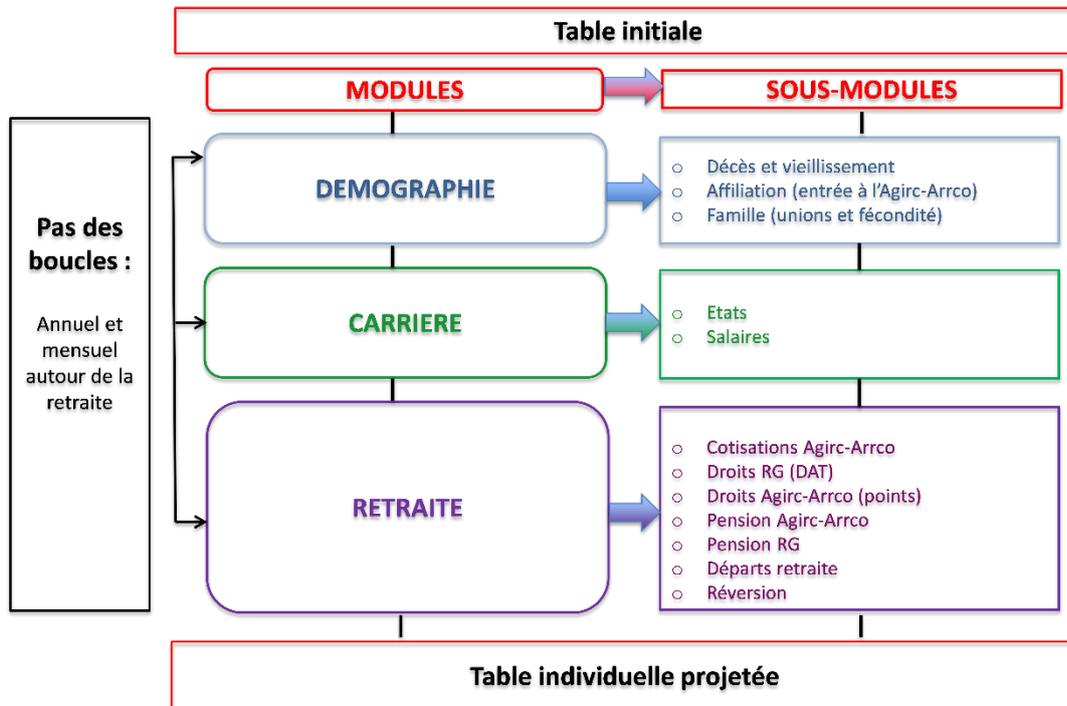
Cependant, la taille sera éventuellement revue en fonction de la qualité des résultats de projection, mais également selon la capacité du serveur et la vitesse de traitement.

Il est également envisagé d'utiliser des échantillons de tailles différentes en fonction du type d'utilisation souhaitée et les contraintes temporelles.

⁹ Il est à préciser qu'une fois qu'un individu est affilié au régime Agirc-Arrco, il apparaît dans le système d'information de l'Agirc-Arrco et il y reste en tant qu'affilié, même s'il ne cotise plus jamais au régime.

3. Architecture du modèles et fonctionnement des modules

a. Architecture du modèle



Ce schéma décrit synthétiquement le fonctionnement du modèle. Il est découpé en différents modules (et sous-modules) qui s'exécutent séquentiellement, cet enchaînement étant répété à chaque pas de temps (annuel) jusqu'à l'atteinte de l'horizon souhaité.

b. Modules

Module démographie

1. Affiliation

L'objectif de ce sous-module est de simuler un flux annuel de nouveaux affiliés à l'Agirc-Arrco et l'initialisation de leurs caractéristiques (démographiques et professionnelles).

A chaque période (année) des individus fictifs sont ajoutés à la base initiale qui constituent les nouveaux affiliés au régime Agirc-Arrco.

A ce stade du développement du modèle, le principe adopté est de maintenir un taux d'affiliation constant en projection, calculé par rapport au nombre de cotisants de l'année précédente. Ce taux d'affiliation est en fonction du sexe, de la naissance en France ou à l'étranger et de l'âge. Il est basé sur la moyenne des taux d'affiliation observés sur les dernières années.

D'autres caractéristiques dont dépend la simulation sont ensuite attribuées à ces nouveaux affiliés : l'âge de fin d'études, le salaire à la première affiliation, les durées d'assurances (validées et cotisées) antérieurement, le nombre d'enfants, etc.

Évidemment, l'hypothèse d'un taux d'affiliation constant est à reconsidérer. Particulièrement, il faudrait tenir compte des hypothèses de la conjoncture du marché du travail, et l'impact des scénarios démographiques (e.g. flux migratoire, fécondité) sur les affiliés du régime.

Famille

Ce sous-module simule la situation familiale (statut marital, nombre d'enfants) dans le but de calculer les droits familiaux (majorations pour enfants et réversions).

Modélisation des naissances :

L'objectif de la modélisation des naissances est d'obtenir le nombre d'enfants de chacun des affiliés de l'Agirc-Arrco.

Cette modélisation détermine à chaque période de la projection la présence d'une naissance ou non, en comparant la probabilité de naissance à un aléa (si l'aléa est inférieur à la probabilité de naissance, une naissance a lieu durant cette période).

Elle est aussi utilisée pour la complétion des données manquantes sur le nombre d'enfants des individus dans la base initiale. Ainsi, elle permet de simuler les trajectoires de fécondité de ces individus afin d'obtenir un stock d'enfants en début de projection (cf. partie 2.1).

La méthode d'estimation est basée sur les salariés du secteur privé provenant du panel DADS – EDP. Les résultats obtenus nous donnent la probabilité d'avoir un enfant supplémentaire en fonction des caractéristiques individuelles (le sexe, l'âge de l'individu, le nombre d'enfants, l'âge des plus jeunes enfants et la position dans la distribution des salaires).

Les naissances obtenues en projection doivent encore être calées afin de respecter l'indicateur conjoncturel de fécondité, selon les scénarios démographiques de l'Insee¹⁰.

Statut marital :

L'objectif est de déterminer, lors de la période de décès, si l'individu a été marié au moins une fois et l'âge du conjoint en cas de mariage. A l'état actuel du modèle, nous supposons que les individus ne se marient qu'une seule fois au cours de leur vie. Ainsi, nous ignorons les remariages et les divorces.

Comme pour la modélisation des naissances, la méthode utilisée pour l'estimation des probabilités de mariage est basée sur les salariés du secteur privé provenant du panel DADS – EDP. Les probabilités obtenues sont en fonction des caractéristiques individuelles (sexe, l'âge de l'individu au moment du décès, la part de la carrière passée dans le secteur privé, la position dans la distribution des salaires). Pour les individus mariés, l'âge du conjoint est prédit par la reproduction de la distribution des écarts d'âge entre conjoints, conditionnellement à ces mêmes caractéristiques et du quartile des écarts d'âge.

2. Décès

¹⁰ A l'heure actuelle, aucun calage n'est prévu pour vérifier que le nombre d'enfants en longitudinal ou en transversal respecte les hypothèses démographiques retenues par l'Insee dans son scénario central.

Le rôle de ce sous-module est de simuler le flux annuel de décédés.

A ce stade, nous appliquons les quotients de mortalité donnés par l'INSEE, l'hypothèse centrale retenue pour le moment étant le scénario de mortalité moyenne.

Ces quotients de mortalité de l'INSEE sont représentatifs de l'ensemble de la population française, donnant ainsi une espérance de vie moyenne selon le sexe et l'âge. Or, plusieurs autres facteurs affectent également l'espérance de vie des individus, notamment le niveau de vie, la catégorie socioprofessionnelle ou le niveau d'éducation. Faire intervenir d'autres déterminants de la mortalité est sans incidence sur le flux annuel de décédés, mais il permet de rendre compte de la répartition de ce flux entre des différentes sous-populations. Cette dimension est particulièrement nécessaire pour la simulation des droits à la retraite de l'Agirc-Arrco. Les ex-cadres (les assurés du régime Agirc) ont en moyenne une meilleure espérance de vie que celle des assurés non-cadres et ils bénéficient en moyenne des droits plus importants. Le fait d'ignorer cette différence d'espérance de vie entre ces deux sous-populations, est susceptible de sous-estimer les dépenses d'allocations du régime. Par conséquent, il est envisagé d'utiliser des tables de mortalité différenciées selon le niveau de points acquis au-delà du PSS.

Module carrière

Ce module simule la carrière des affiliés de l'Agirc-Arrco (les transitions d'états sur le marché du travail et les salaires) à partir de la date initiale (début de la projection) et jusqu'à la cessation définitive d'emploi et/ou de validation de trimestres.

Pour les individus déjà affiliés à la période initiale de projection, le module simule l'évolution de leur carrière. Pour les nouveaux affiliés, la carrière est simulée dès la première entrée au régime.

Cette simulation, conformément à l'ensemble du modèle, suit un pas annuel.

La description ci-dessous offre une idée générale du fonctionnement du module carrière.

1. Transitions d'états sur le marché du travail

Détermination des états :

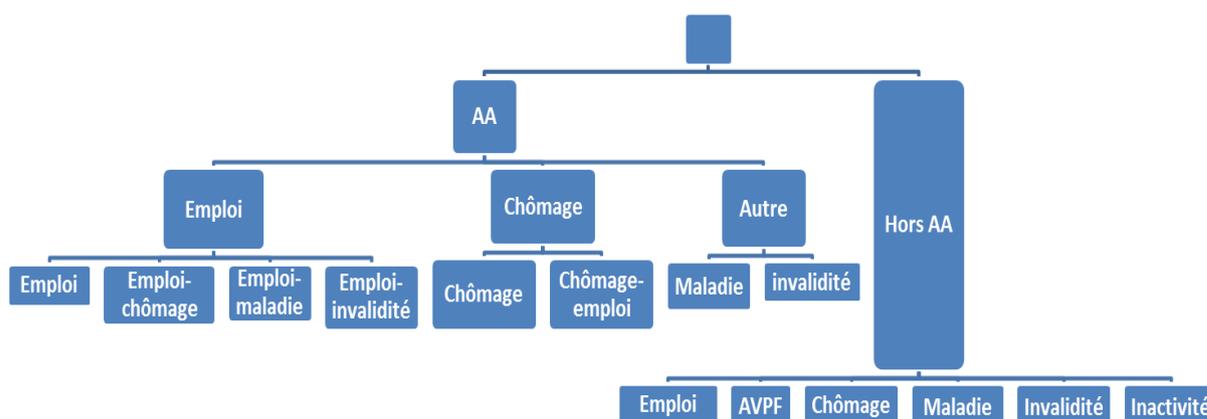
Pour les besoins de la modélisation, un état annuel doit être déterminé à partir de la première année d'affiliation au régime Agirc-Arrco (première acquisition de points en emploi et premier salaire perçu). Les périodes antérieures à l'affiliation ne sont pas modélisées. Cependant, et bien qu'il s'agisse d'un modèle « régime », les périodes en dehors du régime sont modélisées à partir de la première sortie du régime. Ainsi, nous distinguons deux types d'états. Les états pour les périodes d'affiliation à l'Agirc-Arrco et les états passés en dehors de l'Agirc-Arrco.

Pour les périodes où l'individu est affilié à l'Agirc-Arrco, l'état annuel est déterminé en fonction de la part du type des points acquis dans les points totaux de l'année. Un état est défini en emploi (chômage) si les points acquis en emploi (au chômage) représentent au moins 95 % des points de l'année. Ainsi, des états simples dominants sont déterminés. Les états doubles sont constitués d'un

état dominant (proportion de points plus élevée) et d'un état secondaire : un individu est en état emploi-chômage, l'emploi étant son état dominant de l'année, si les points acquis sur l'année sont constitués des points acquis via l'emploi et le chômage, et que les points acquis via le chômage représentent entre 5% et 50 % des points totaux. Inversement, si les points chômage sont compris entre 50 % et 95%, l'état dominant de l'année est le chômage.

La figure suivante donne des exemples de types d'états retenus, qui diffèrent légèrement selon les tranches d'âge : emploi¹¹, emploi-chômage, emploi-maladie, emploi-invalidité, chômage, chômage-emploi, maladie, invalidité.

Les états pour les périodes passées en dehors de l'Agirc-Arrco (CNAV + MSA + autres régimes), sont déterminés en fonction des trimestres validés dans les autres régimes. Le type de trimestre qui représente le maximum de trimestres validés dans tous les régimes définit un seul état annuel : emploi, avpf, chômage, maladie, invalidité, inactivité (hors retraités). Les mêmes états sont utilisés en projection.



¹¹ A ce stade, la quotité de travail annuelle n'est considérée qu'à partir du nombre de jours en emploi - nous faisons l'hypothèse de temps plein pour un jour de travail. De même, le cumul emploi-retraite n'est pas modélisé.

La simulation des transitions d'états sur le marché du travail nécessite de modéliser les transitions individuelles entre états en amont de l'exercice de projection. Cette modélisation est réalisée sur la moitié des individus de l'échantillon sélectionné, que nous appelons « échantillon d'apprentissage ». Les estimations sont réalisées par tranches d'âge (début, milieu et fin de carrière). Le modèle utilisé dans les estimations est le "Logit imbriqué". Le principe de ce modèle est d'estimer des probabilités d'appartenance aux différents états d'une façon séquentielle, du niveau général au niveau le plus fin (cf. figure ci-dessus). Un premier niveau estime la probabilité d'affiliation au régime Agirc-Arrco. Conditionnellement à l'affiliation au régime, un second niveau estime les probabilités d'être en emploi, au chômage ou à un autre état (maladie, invalidité). Ensuite, pour chacun de ces trois états, un dernier niveau estime les probabilités d'appartenance aux états finals (e.g. pour le chômage, la probabilité d'être dans un état chômage simple ou un état double chômage-emploi). Dans le cas des années passées en dehors du régime, les probabilités d'être en emploi, avpf, chômage, maladie, invalidité ou inactivité sont estimées.

Les variables de contrôles retenues sont celles qui offrent une meilleure qualité de prédiction au regard des transitions d'états réellement observées. Ces variables ne sont pas exactement identiques entre les différents niveaux d'équations estimées, et elles varient également selon les tranches d'âge. En général, il s'agit de l'âge et de son carré, le type d'état de l'année précédente et des variables indiquant le nombre d'années accumulées dans chacun des états depuis l'affiliation au régime. L'état de l'année précédente et les variables de cumuls des différents états améliorent considérablement la qualité de prédiction du modèle. Intuitivement, le type d'état de l'année précédente affecte la probabilité de transition vers un tel ou tel type d'état. D'une manière générale, le fait d'être dans un type d'état donné l'année en cours augmente la probabilité de transiter vers ce même type d'état l'année suivante. Les cumuls des états permettent de rendre compte de l'impact de l'historique de la carrière, et notamment de la permanence dans les états.

Pour chaque année de projection et jusqu'à le départ à la retraite (cf. module retraite – départ en retraite), ce processus d'équations séquentielles sert à attribuer un état annuel pour chaque individu simulé. Un individu inactif en fin de sa carrière, mais qui n'a pas encore atteint les conditions requises pour partir au taux plein (hypothèse retenue à ce stade pour le comportement de départ), rentrerait dans cet enchaînement séquentiel et un état lui est attribué.

Afin de confirmer la validité des estimations des équations de transitions, quelques tests sont ensuite réalisés. Les trajectoires année par année sont simulées sur l'autre moitié de l'échantillon (« échantillon test ») à l'aide du processus d'équations séquentielles et des coefficients obtenus de leur estimation. Ensuite, nous comparons les trajectoires observées et prédites de l'échantillon test : probabilité d'être dans les différents états par âge, distribution du nombre total d'années passées dans les différents états, distribution du nombre de transitions sur la période concernée, probabilité de transition par âge, probabilité de transition par âge à chaque état ($t+1$) conditionnellement à l'état (t).

2. Salaire

Comme pour les transitions d'états, une modélisation des salaires est effectuée au préalable. Les coefficients qui en résultent servent à la projection des salaires.

Les salaires utilisés dans la modélisation sont des salaires reconstitués à partir des points par tranche acquis en emploi et des taux de cotisation correspondants. Les estimations sont réalisées à la fois sur le salaire annuel effectivement perçu et sur le salaire annualisé (en équivalent temps plein). Le pas temporel d'estimation est un pas annuel et la période de référence utilisée pour estimer les équations va de 1980 à 2012. Les équations sont estimées séparément pour le début de carrière (avant l'âge de fin d'études) et pour la carrière principale, ainsi que selon le sexe. Pour chacun de ces groupes deux estimations sont réalisées : la première estimation concerne les individus ayant travaillé dans le régime Agirc-Arrco pour au moins une année durant les cinq dernières années. La seconde estimation pour ceux ayant une interruption de carrière dans le régime Agirc-Arrco qui remonte à cinq ans au moins.

La modélisation se fait en deux étapes : 1) des équations de salaires sont estimées au niveau de l'individu-année, en régressant le salaire sur un ensemble de variables explicatives. Ces estimations permettent d'obtenir des valeurs de référence (les coefficients) aux paramètres des équations de salaires, ainsi que des termes résiduels donnant une mesure des hétérogénéités inobservables au niveau de l'individu-année, qui ne sont pas contrôlées par les différentes variables utilisées dans la modélisation. En projection, les valeurs de référence sont utilisées afin d'obtenir pour chaque individu simulé une prédiction de la partie déterministe du salaire. 2) Estimation d'équations des carrés des résidus individuels obtenus de la première étape : les carrés des résidus individuels sont régressés sur quasiment les mêmes variables de contrôles de la première étape, afin d'obtenir en projection une estimation des écarts-types individuels qui vient s'ajouter à la partie déterministe du salaire et ajouter de la dispersion au sein des catégories d'individus ayant les mêmes caractéristiques observées.

Les variables de contrôles utilisées dans les modélisations sont : la durée (nombre d'années) en emploi en Agirc-Arrco et son carré, la durée en emploi en Agirc-Arrco sans interruption et son carré, l'âge de fin d'études (âge d'insertion sur le marché du travail), interaction des variables des durées en emploi et leurs carrés avec l'âge de fin d'études par tranches d'âge, indicatrice pour la première année de carrière en AA, indicatrice pour la naissance à l'étranger, indicatrice pour au moins une année où aucun points emploi n'a été validé au cours des 5 années précédentes et l'interaction de cette variable avec l'âge de fin d'études, indicatrice pour au moins un point chômage validé durant l'année, indicatrice pour au moins un point maladie validé durant l'année, salaire de l'année précédente, 0 sinon et considérer le salaire d'il y a 2 ans, 0 sinon et considérer le salaire d'il y a 3 ans, ainsi de suite jusqu'à le salaire d'il y a 5 ans.

Ces différentes variables et la spécification retenues sont celles qui expliquent le mieux le salaire, relativement à d'autres spécifications utilisées. Les variables de la durée en emploi donnent une mesure de l'expérience professionnelle de l'individu qui affecte le niveau de son salaire actuel. Le fait d'inclure également la durée sans interruption permet de rendre compte de l'impact de l'ancienneté/continuité dans la carrière, les individus ayant des carrières heurtées ont souvent des salaires plus faibles. L'âge de fin d'études influence la trajectoire salariale de l'individu, en étant une approximation de son niveau d'éducation ou du commencement d'une carrière stable. Le dernier salaire obtenu permet de capturer une évolution salariale. La spécification quadratique des variables donne plus de flexibilité (une non linéarité) aux effets estimés. De même, un impact différencié selon l'âge de fin d'études est pris en compte par les variables d'interactions.

Les principaux tests de validation des estimations effectués :

1. Validation de la qualité de la spécification retenue : une estimation est faite sur la moitié des individus de l'échantillon sélectionnée « échantillon d'apprentissage » et la simulation des salaires est effectuée sur l'autre moitié « échantillon test », sur une année donnée. Les distributions des salaires prédits et observés sont ensuite comparées.
2. Validation de la qualité de la simulation : une simulation des salaires année par année est effectuée sur un champ plus large d'individus, qui exclut l'échantillon d'apprentissage, mais qui ne respecte pas les critères de l'échantillon sélectionnée pour les estimations. Ensuite, une comparaison des distributions des masses salariales individuelles prédites et observées sont comparées.

Il n'y a pas d'autres revenus projetés à ce stade.

Module retraite

1. Cotisations

L'objectif de ce sous-module est de calculer les cotisations relevant du régime Agirc-Arrco.

Les cotisations d'emplois et de chômage sont calculées par type de tranches, et en considérant la législation correspondant à la période antérieure au régime unifié et celle correspondant au régime unifié.

2. Droits

Ce sous-module calcule les droits acquis auprès du régime Agirc-Arrco, mais également ceux acquis auprès du régime général.

Tous les types de points Agirc-Arrco sont calculés par tranches et selon la législation en cours de la période d'acquisition des points - antérieure au régime unifié et depuis le régime unifié.

A partir de l'état sur le marché du travail déterminé par le module carrière, les trimestres acquis au régime général et dans les autres régimes sont définis selon un tirage dans des distributions empiriques en fonction du sexe, de la naissance en France ou à l'étranger, et de l'âge.

Ces droits inter-régimes sont nécessaires parce qu'ils constituent la durée d'assurance tous régimes et en conséquence ont une incidence sur les décisions de départ.

3. Départ en retraite

L'objectif de ce sous-module est de simuler les comportements de départ à la retraite des affiliés de l'Agirc-Arrco.

A ce stade, une version simplifiée des comportements de départ, qui est basée sur la recherche du taux plein, est simulée par le modèle. Cette version tient compte de la retraite anticipée pour carrière longue.

Une modélisation plus fine des comportements de départ est en développement. Même si le pas temporel du modèle est un pas annuel, la modélisation se fait sur un pas mensuel l'année de départ à la retraite, notamment pour tenir compte de la montée en charge de la réforme de 2010 pour l'AAD et éventuellement des réformes suivantes mais aussi des besoins d'informations plus précises pour les prévisions d'activités de liquidation en gestion.

4. Pension

Ce sous-module sert à calculer le montant de la pension Agirc-Arrco par type de tranches, en tenant compte du changement de la législation à la suite du régime unifié. A ce stade, la pension du régime général n'est pas encore calculée.

5. Réversion

A ce stade du développement du modèle, ce sont uniquement les pensions de réversion pour conjoints qui sont considérées en simulation¹². Les mariages multiples des assurés qui ouvrent des droits dérivés, et le remariage de leurs conjoints survivants ne sont pas pris en compte.

4. Calages

Dans la plupart des situations, l'étape du calage intervient uniquement dans un second temps et ne contribue pas à la construction du modèle. En principe, il est souhaité de ne procéder au calage qu'après avoir testé le comportement du modèle non calé. Néanmoins, c'est une étape importante qui permet à la simulation de respecter différents scénarios (démographiques, macroéconomiques), ou de s'aligner sur des cibles déterminées d'une façon exogène au modèle.

1. Effectifs de cotisants et de chômeurs Agirc-Arrco :

Le calage des effectifs de cotisants et de chômeurs est essentiel, il dépend d'un ensemble d'hypothèses économiques (le taux de chômage, le taux d'emploi), démographiques (fécondité, mortalité, migration) et de l'âge de départ à la retraite. Il est ainsi le résultat de l'interaction entre plusieurs modules du modèle.

Les effectifs de cotisants et de chômeurs Agirc-Arrco doivent respecter les cibles de salariés du secteur privé et du taux de chômage, qui sont fournies par le COR ou définies en interne par l'Agirc-Arrco dans le cadre des négociations paritaires.

La variation annuelle du stock de cotisants dépend des flux des entrants de nouveaux affiliés, des retours au régime pour ceux qui ont été déjà affiliés, des flux sortants du régime, des retraités et des décès. La cale est appliquée dans le module carrière, en ajustant le stock de cotisants par les flux entrants de retour dans le régime (hors première affiliation), et les flux sortants du régime (hors retraites et décès).

2. Salaire moyen par tête :

¹² Les pensions de réversion attribuées aux orphelins (représentant une fraction minimale) ne sont pas implémentées.

L'évolution du SMPT doit respecter les différentes hypothèses fournies par le COR ou définies en interne dans le cadre des négociations paritaires, sur toute la période (2019-2070).

Comme précisé en introduction, le développement du modèle de l'Agirc-Arrco n'est pas totalement achevé, il sera mis en production au cours de l'année 2020. Cette année sera consacrée aux tests d'évaluation du modèle, à la question des calages, et à l'amélioration de certains modules (notamment le choix du départ en retraite).