

CONSEIL D'ORIENTATION DES RETRAITES

Séance plénière du 25 janvier 2012 à 9 h 30

« Exercices de projections financières du système de retraite : aspects méthodologiques »

Document N°5

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

Quelques modèles de projection en matière de retraite à l'étranger

Secrétariat général du Conseil d'orientation des retraites

Quelques modèles de projection en matière de retraite à l'étranger

Ce document présente les modèles de projection en matière de retraite utilisés dans quatre pays étrangers choisis dans le souci de présenter une diversité de méthodes de projection: l'Allemagne, l'Italie, la Suède et les Etats-Unis. Il précise le **document n° 6** du dossier qui porte plus largement sur les exercices de projection à l'étranger.

Au sein de l'Union Européenne, le Comité de Politique Economique (CPE) a mis en place un groupe de travail sur les effets du vieillissement, dit *Ageing Working Group* (AWG), (cf. **document n°7**). Dans ce cadre, les Etats-membres réalisent des projections à long terme en matière de retraite et décrivent les méthodes et hypothèses retenues pour leurs modèles de projection. Les informations relatives à l'Allemagne, l'Italie et la Suède sont notamment issues des travaux passés de ce groupe.

Pour les Etats-Unis, deux modèles sont présentés, celui de la SSA et du CBO.

1. Allemagne : deux modèles à générations imbriquées couplés

1) Présentation générale

Le modèle allemand de projection financière à long terme du régime légal de retraite est géré conjointement par le Ministère des affaires sociales et l'assurance vieillesse. Il est utilisé pour la publication annuelle d'un rapport sur la soutenabilité budgétaire à un horizon de 15 ans.

Il est bâti à partir de deux sous-modèles différents.

Le premier est un modèle par génération qui permet de déterminer l'impact des évolutions démographiques sur le montant des pensions. Le second a vocation à présenter l'évolution financière selon différents indicateurs : taux de cotisation, ajustement des pensions...

Ces modèles utilisent deux modules différents :

- l'un projetant les carrières ;
- l'autre modélisant les âges de départ à la retraite.

Les données utilisées sont pour la plupart issues des travaux de l'office de statistique et de démographie allemand. Les données principales concernent le nombre de pensionnés et la pension moyenne, en stock et en flux.

Les hypothèses démographiques et macroéconomiques (population, mortalité, marché du travail, salaires) sont celles définies pour les projections de l'AWG.

2) Architecture

a) *Le modèle démographique*

Dans un premier temps, le modèle démographique détermine par génération l'évolution annuelle du nombre de retraités et du montant moyen des pensions.

Le nombre de retraités évolue du fait de l'arrivée de nouveaux retraités, dont les départs en retraite sont projetés par le module correspondant, de la mortalité des personnes déjà à la retraite et des éventuels bénéficiaires de pensions de réversion dues à ces décès.

Le montant moyen des pensions est obtenu en pondérant les pensions des retraités déjà présents et celles des nouveaux entrants (effet noria). Le montant moyen de pension des nouveaux retraités est estimé grâce au module de projection des carrières, celui des retraités déjà présents est corrigé des décès et des éventuelles réversions.

Les dépenses de pension annuelles se déduisent des prévisions annuelles d'effectifs de pensionnés et de montant moyen des pensions. Toutefois, cette approche ne met en lumière que les tendances démographiques et celles du marché du travail, et pas la dynamique financière du régime.

b) *Le modèle financier*

Dans une seconde étape, le modèle a vocation à présenter la dynamique des dépenses de pension et des recettes du régime. Il utilise les résultats obtenus à l'issue de la première étape, en considérant de surcroît l'indexation des dépenses de pension, plus précisément la revalorisation du point. En raison de la contrainte légale d'équilibre budgétaire, le taux de cotisation s'en déduit également.

La valeur du point est revalorisée chaque année en prenant en compte :

- l'évolution du ratio de dépendance « nombre de pensionnés / nombre de cotisants » ;
- l'évolution du taux de cotisation ;
- la progression du salaire moyen.

Plus précisément, pour le calcul du ratio de dépendance, le nombre de pensionnés (respectivement de cotisants) est déterminé en « équivalent moyen », c'est-à-dire que le dénombrement est effectué en calculant le rapport entre le total des prestations versées (respectivement des cotisations reçues) et la pension standard¹ (respectivement la cotisation standard versée par un assuré au salaire moyen).

Les recettes du régime légal de retraite sont constituées de l'ensemble des cotisations versées et des transferts budgétaires. Les cotisations sont fonction de l'évolution des salaires, du nombre de personnes en emploi, mais aussi, par exemple, du nombre des chômeurs dans la mesure où l'agence fédérale pour l'emploi transfère des cotisations pour maintenir les droits à pension. *In fine*, le taux de cotisation est fixé pour garantir l'équilibre du régime et constitue l'un des indicateurs retenus.

¹ Pension perçue par un retraité au terme d'une carrière de 45 ans au salaire moyen.

2. Italie : un modèle « multi-états »

1) Présentation générale

Le modèle italien est un modèle dit *multi-états*, mis à jour annuellement depuis 1999. Il est sous la responsabilité du Ministère des finances (*Ragioneria Generale dello Stato*), dans le cadre d'un exercice obligatoire annuel de projection.

Le système italien comporte de nombreux régimes, dont les données - fournies par les principaux organismes : INPS pour les régimes privés et INPDAP pour les fonctionnaires - sont centralisées au sein du Ministère des finances. Il est à noter que l'INPS dispose de son propre modèle de microsimulation, mais dont les résultats ne sont pas publiés.

Les hypothèses démographiques sont issues des projections officielles réalisées par l'institut de statistique et démographie italien (ISTAT). Le Ministère des finances fournit quant à lui les hypothèses économiques.

2) Architecture

Le modèle est composé de quatre modules différents : la démographie, le marché du travail, la productivité et les pensions. Ces modules interagissent les uns avec les autres. Le module central, celui relatif aux pensions, reçoit notamment les données fournies par les trois autres.

Les modules retiennent une **approche par génération**.

1. Le module démographique réalise des projections de population, par âge et par sexe, selon les tables de mortalité et les prévisions de naissances et de migrations. Les résultats de ce module fournissent des données relatives à l'offre de travail utiles au module marché du travail.

2. Le module « marché du travail » projette la population en emploi en considérant les évolutions tendanciennes du marché du travail. Il fonctionne en interaction avec le module « pensions ». En effet, outre l'allongement de la durée des études, l'offre sur le marché du travail est également affectée par les évolutions de la législation en matière d'ouverture des droits à retraite. Le taux de chômage est calculé par âge et par sexe, et vérifie une hypothèse de convergence autour d'une valeur moyenne, à une vitesse qui dépend de l'évolution de la population active. Le nombre d'heures travaillées, dépendant de la part d'agents à temps partiel et de la durée moyenne de travail, est également déterminé pour alimenter le module « productivité ».

3. Le module « productivité » prend comme hypothèse exogène la productivité totale des facteurs (supposée constante à long terme) et considère les modifications d'intensité capitalistique. Une fonction de production Cobb-Douglas est utilisée.

4. Le module « pensions » vise à décrire le parcours de chaque individu, depuis le paiement de sa première cotisation jusqu'à son décès. Il utilise des données fournies par chacun des modules précédents.

- a) le module démographique détermine des probabilités de décès affectant la population pensionnée ;
- b) le module « marché du travail » modélise les nouveaux entrants (premières cotisations) ;
- c) le module « productivité » fournit des données sur l'évolution des salaires.

Pour chaque individu représentatif, à chaque année est associée une situation, qui est le croisement de plusieurs variables d'état : le statut (retraité, cotisant, en « sommeil », ou retraité et cotisant la même année), le sexe, l'âge, le type de retraite, la durée de cotisation...

Le modèle repose sur une matrice contenant les probabilités de transition entre les états, ainsi que sur des probabilités de survie, d'où la dénomination de modèle *multi-états*.

Une valeur monétaire moyenne (salaire ou pension) est associée à chaque état et est actualisée après chaque transition. Les nouvelles pensions sont déterminées à partir des probabilités d'arriver à l'état de retraité, une fois les conditions d'âge et de cotisation validées. Le montant de la pension est calculé à partir des données monétaires issues du parcours décrit précédemment, selon la réglementation en vigueur.

3. Suède : un modèle de microsimulation

Deux modèles de microsimulation sont utilisés en Suède afin de réaliser les exercices de projections en matière de retraite :

- l'un, le modèle *Swedish MicroSimulation* (SESIM), est géré par le Ministère des finances suédois en coopération avec des chercheurs suédois et utilisé dans le cadre des projections pour l'AWG;
- l'autre, le modèle MIMESIS, géré par l'Agence de sécurité sociale, est utilisé pour les exercices nationaux (publication du *Orange report*).

Seul le modèle SESIM est décrit ici.

1) Présentation générale

Le modèle SESIM repose sur la base de données initiale *Longitudinal Individual Data* (LINDA), qui comprend 3,5 % de la population suédoise. Les données sont issues du croisement de nombreuses sources administratives (et non d'enquêtes) relatives au salaire, à l'état de santé, aux droits à pension... Les individus sélectionnés sont suivis dans le temps et les informations sont disponibles depuis 1960.

Cette base de 110 000 individus simule, à un pas annuel, différents événements de la vie et de la carrière d'un individu : formation, mariage, naissance, travail, retraite... Le modèle retient principalement une approche individuelle, mais SESIM considère également le ménage comme unité d'observation.

SESIM exploite un sous-échantillon de LINDA, calé sur données macroéconomiques.

Deux ajustements sont effectués :

- le premier est relatif à la notion de foyer. Comme LINDA ne considère pas le foyer au sens économique, des retraitements et imputations sont nécessaires ;
- le second concerne les individus ayant des droits à pension mais ayant émigré et, de ce fait, disparu des registres administratifs suédois. Une partie d'entre eux est réintégrée dans le modèle pour corriger cette sous-estimation.

2) Architecture

Le modèle considère les salaires et les distributions d'âge sur le marché du travail comme des variables endogènes. Chaque année, un statut est associé à chaque individu parmi les suivants : enfant, étudiant, chômeur, employé, émigré, en congé parental, retraité, invalide, divers.

La microsimulation est effectuée selon plusieurs étapes successives, à un pas annuel.

1. La première étape est constituée d'un **module démographique** simulant des événements tels que les naissances, les migrations, les mariages, les séparations et les décès.

2. L'étape suivante représente les éventuels changements relatifs à **l'éducation** de l'individu : passage d'un niveau d'enseignement à un autre, sortie de l'enseignement, formation continue...

3. Un module relatif au **marché du travail** permet ensuite de simuler les périodes d'emploi (supposée nécessairement à temps plein) et de chômage. Les revenus sont également simulés dans cette étape ; les données de salaires sont par exemple estimées par des équations.

Concernant le choix de départ à la retraite, des décisions individuelles peuvent être modélisées, mais les calculs retenus pour l'AWG reposent sur l'hypothèse de départs systématiques à 65 ans.

4. L'étape suivante simule les **autres sources de revenus** de l'individu : revenus financiers, revenus du capital. Les règles relatives aux trois piliers du système de retraite sont prises en compte : retraite publique, pension issue de l'activité privée, et épargne privée.

5. Des éléments de **redistribution** sont ensuite simulés, à travers la prise en compte de la fiscalité et de certaines allocations.

4. Etats-Unis : deux modèles, dont un modélisant l'incertitude

Deux modèles de projection sont présentés ici : celui de la Social Security Administration (SSA), et celui du *Congressional Budget Office* (CBO). Ils couvrent le champ du système de Sécurité sociale fédérale correspondant au programme *Old-Age, Survivors, and Disability Insurance* (OASDI), dont bénéficient chaque année environ 56 millions d'individus.

Ces bénéficiaires se répartissent comme suit :

- 69% de retraités de droit direct ;
- 12% de retraités au titre de la réversion ;
- 19% de travailleurs invalides ou leurs proches.

Deux fonds différents sont dédiés chacun à un volet de la protection sociale : le premier, *Old Age Survivors Insurance* (OASI), couvrant le risque-vieillesse (droits direct et dérivé), le second, *Disability insurance* (DI), couvrant le risque-invalidité.

1) Le modèle de la *Social Security Administration* (SSA)

a) Hypothèses et indicateurs

Aux Etats-Unis, l'exercice de projection du système de retraite obligatoire est réalisé par l'actuaire en chef de la sécurité sociale. Il donne lieu à la publication annuelle d'un rapport, dont l'horizon temporel est de 75 ans. Ce rapport a vocation à apprécier la situation financière du régime.

Le modèle actuariel repose sur des hypothèses fixées par les actuaires de la sécurité sociale et est établi à partir d'un scénario central. Deux autres scénarii sont également envisagés, l'un pessimiste, l'autre optimiste.

L'indicateur principalement retenu par la SSA est l'année d'épuisement des réserves (Social Security Trust fund).

b) Un modèle actuariel agrégé

Le modèle de la SSA est un modèle actuariel agrégé, ou en cellules, dit « *cells-based model* ». La population est répartie en agrégats (par sexe, âge, statut marital) et projetée selon des paramètres déterministes. L'environnement macroéconomique est exogène.

Trois composantes principales du modèle peuvent être distinguées : un module démographique, un autre de prévisions à court terme, mis à jour tous les trimestres, et un dernier de prévisions à long terme, actualisé annuellement.

La population est projetée en considérant des événements tels que les naissances, décès, mariages, divorces, veuvages et les flux migratoires nets. Le modèle considère les transitions d'année en année entre les différents statuts suivants : célibataire, marié, divorcé, veuf.

Les recettes sont également projetées, elles sont en très grande partie liées aux cotisations notamment, qui dépendent du nombre de personnes en activité, le reste étant composé des intérêts perçus sur les réserves.

Les dépenses correspondent quant à elle principalement aux risques vieillesse et invalidité et sont plus largement fonctions de la distribution de la population totale, par âge, sexe et statut.

Pour le module de court terme comme pour celui de long terme, les projections sont effectuées en deux étapes successives :

- la première consacrée à la projection des emplois et des revenus pour déterminer les recettes ;
- la seconde aux dépenses.

Dans le cas du module de court terme, le modèle considère des hypothèses de PIB et d'inflation. Le taux d'activité et le chômage, estimés par méthodes économétriques (par âge, sexe, année) permettent de déterminer le nombre total de travailleurs. Cette quantité est ensuite répartie entre activité salariée et activité non salariée. Les revenus totaux puis, plus précisément, ceux des individus couverts par l'OASDI sont projetés par secteur économique. Finalement, les cotisations sont déterminées en considérant les cotisations patronales, les cotisations salariales et les cotisations des non-salariés (chacune étant soumise à un taux qui lui est propre).

L'étape suivante consistant à estimer les dépenses est effectuée en projetant tout d'abord le nombre de bénéficiaires, puis le montant moyen perçu par individu. Les taux de départ sont estimés à partir des tendances historiques, en incluant les modifications législatives.

Les modules de court terme et de long terme diffèrent dans les modèles économétriques retenus pour représenter les taux d'activité et de chômage.

Cette démarche repose sur une approche **déterministe**, les valeurs en entrée du modèle prenant des valeurs moyennes fixes. Une autre approche, dite **stochastique**, a été retenue par le *Congressional Budget Office* (CBO), rattaché au Congrès des Etats-Unis.

2) Le modèle du *Congressional Budget Office* (CBO)

Le modèle de projection du CBO est un modèle de microsimulation individuelle.

Outre les années auxquelles ces réserves seront épuisées, le modèle de projection retient notamment comme indicateurs les recettes, les dépenses et le déficit en pourcentage de PIB. En plus d'un scénario conforme aux dispositions législatives en vigueur, le modèle prévoit également des scénarii intégrant des modifications réglementaires.

Les dépenses de retraite sont estimées et restituées sous la forme de distribution (quintiles et médiane pour chaque cohorte tous les 10 ans : individus nés dans les années 1940, ceux nés dans les années 1950...). Une ventilation par sexe est prévue. Les montants de pension sont calculés sous l'hypothèse de liquidation à 65 ans. Les engagements financiers et des éléments de distribution sur les taux de remplacement sont également publiés.

Le modèle du CBO réalise des projections à l'horizon de 75 ans. A l'instar des modèles classiques de microsimulation, il simule les événements démographiques tels que les décès, naissances, unions, mouvements migratoires et les données relatives au marché du travail (participation, nombre d'heures travaillées...).

Sa particularité est d'introduire des éléments de dispersion et de pouvoir ainsi mesurer la part d'incertitude des résultats des projections, sous la forme d'intervalles de confiance. En cela, il s'agit d'un modèle **stochastique**.

a) Les données et hypothèses utilisées par le CBO

Le modèle du CBO repose sur un échantillon représentatif de la population américaine.

Les principales données sont issues des fichiers de la Sécurité sociale (*Continuous Work History Sample*). D'autres informations sont imputées et calées sur certaines données issues principalement d'enquêtes : enquêtes revenu et emploi (*Survey of Income and Program Participation, SIPP*), données de panel sur les revenus (*Panel Study of Income Dynamics, PSID*) et enquêtes démographiques (*Current Population Survey, CPS*). Par exemple, les unions sont imputées de manière à respecter les proximités et corrélations entre partenaires sur des critères tels que l'âge, le salaire et le niveau d'éducation. Le taux de mortalité est affiné selon le revenu.

Concernant la participation au marché du travail, le modèle impute pour chaque individu des informations telles que la quotité de travail, le nombre d'heures travaillées, le revenu annuel et les périodes de chômage. Plus précisément, il détermine des probabilités à partir de variables explicatives comme l'âge, le sexe, le niveau d'éducation, la situation maritale et le nombre d'enfants en bas âge. Les salaires moyens sont par ailleurs estimés à partir d'équation ayant pour variables l'âge, le genre et le niveau d'éducation.

b) Une approche stochastique

Contrairement aux modèles de microsimulation classiques retenant les valeurs moyennes de chaque variable, celui du CBO inclut une approche stochastique par simulation de Monte-Carlo.

Les méthodes de Monte-Carlo (voir plus loin pour une description technique plus précise) consistent à calculer une valeur numérique à partir de données aléatoires, dont la distribution probabiliste est connue. La valeur est calculée en itérant un grand nombre de fois la simulation conformément à la distribution de la variable.

Contrairement à la logique déterministe qui prévaut dans la plupart des exercices de simulation, cette approche permet d'introduire de la variabilité dans les résultats des projections et d'en déduire des intervalles de confiance.

C'est ainsi que le salaire individuel comprend, outre la moyenne de la catégorie retenue, un choc aléatoire introduisant de la dispersion.

c) Le bouclage macroéconomique

Le bouclage macroéconomique détermine le PIB à partir de la productivité, de la quantité de travail et du stock de capital productif. Le taux de croissance de la productivité totale des facteurs est considéré exogène. Le nombre d'heures travaillées, quant à lui, est fonction de la force de travail et du taux de chômage.

Les hypothèses économiques reposent sur les analyses du CBO.

d) La modélisation de l'incertitude affectant les hypothèses

La méthode de Monte-Carlo retenue consiste à modéliser chacune des variables servant d'hypothèse en entrée du modèle et à répéter de nombreuses fois l'exercice de projection sur la base d'hypothèses différentes tirées aléatoirement. Les hypothèses sont d'ordre démographique (taux de fécondité, allongement de l'espérance de vie, immigration nette, évolution de la santé) et économique (inflation, chômage,...).

Les variables sont modélisées à partir des observations historiques sous forme de séries temporelles. D'une date t à la suivante $t+1$, la valeur de chaque variable est composée d'une partie prévisible (à partir des observations antérieures) et d'une partie aléatoire qui permettra d'introduire de la dispersion.

Plus précisément, une variable en entrée du modèle x_t , à un instant t , peut être de manière très simplifiée décomposée comme la somme de trois termes :

$$x_t = \alpha + \beta x_{t-1} + \varepsilon_t \quad (\text{Modèle autorégressif})$$

- le premier terme, α , représente la tendance centrale ;
- le deuxième, β , mesure la persistance dans le temps (corrélacion entre deux valeurs successives) ;
- le dernier terme, ε_t , permet d'estimer la variance du processus de choc qui capte la variabilité.

D'autres modélisations peuvent être retenues (avec des chocs corrélés ou bien encore en introduisant davantage de persistance).

La variance du dernier terme, estimée, permet de calculer des intervalles de confiance qui s'élargissent au fur et à mesure que l'horizon de prévision est lointain.

Deux types de modèle sont retenus pour les variables démographiques

- une modélisation AR(1) pour l'espérance de vie par exemple ;
- une modélisation ARMA(4,1) pour le taux de fécondité ou l'immigration nette.

Concernant les variables économiques, deux cas peuvent se présenter :

- certaines variables sont modélisées indépendamment les unes des autres. Tel est le cas de la productivité totale des facteurs, modélisée par un bruit blanc (i.e. : $\alpha = \beta = 0$), ou bien encore la part des revenus en capitalisation, modélisée en AR(1) ;
- d'autres évoluent conjointement, impliquant une modélisation sous forme de vecteur autorégressif.