

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

Document N° 5

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

Description du modèle Destinie

Insee, Direction des Études et Synthèses Économiques

Description du modèle Destinie

1. Présentation générale du modèle

Le modèle Destinie (modèle Démographique Économique et Social de Trajectoires Individuelles simulées) est un modèle de microsimulation dynamique développé et géré par l'Insee, dont l'objectif principal est la projection à long terme des retraites. Il a été mobilisé, depuis le milieu des années 1990, pour un grand nombre d'analyses publiées par l'Insee. Sa version actuelle (« Destinie 2 ») a été développée dans la seconde moitié des années 2000. Son cœur de calcul est désormais codé en C++ et une librairie R permet de lancer des simulations. Le code source du modèle est librement accessible en ligne dans son intégralité sur la plateforme github : <https://github.com/InseeFr/Destinie-2>.

L'approche de Destinie consiste à simuler le devenir d'un ensemble d'individus constituant un échantillon représentatif de la population française. Cet échantillon est issu des données de l'enquête Patrimoine de l'Insee ; son champ est celui des personnes résidant en France, vivantes en 2010 ou nées après cette date. L'outil modélise, pour toutes les personnes, un ensemble d'événements démographiques (mises en couple, naissances, séparations, migrations et décès) et professionnels (transitions entre périodes d'activité et d'inactivité, entre statuts d'emploi, évolution des revenus salariaux) ainsi que les décisions individuelles de départ à la retraite, et il calcule le montant des pensions en appliquant les barèmes prévus par la réglementation. Les principaux régimes de retraite français sont modélisés : le régime général (regroupant tous les salariés du secteur privé et les contractuels du secteur public), les régimes complémentaires Agirc et Arrco, le régime de la fonction publique (regroupant tous les fonctionnaires), ainsi que le régime des indépendants (modélisé comme le régime de base du SSI).

La construction de l'échantillon démographique donne au modèle une perspective plus généraliste : celle d'un outil de simulation des trajectoires à long terme des individus et de leurs ménages. Celle-ci permet de répondre à des questions directement liées au vieillissement, ou nécessitant la disponibilité de données de panel en très longue période, combinant une dimension rétrospective et une dimension prospective. Sa dimension « ménages » le distingue d'autres modèles de microsimulation : les liens familiaux étant simulés et actualisés chaque année de projection, Destinie permet de calculer les pensions de réversion et d'évaluer les effets budgétaires et redistributifs d'une modification de la réglementation concernant les droits dérivés.

La taille relativement modeste de l'échantillon de Destinie (environ 62 000 personnes) et sa programmation dans un langage de « bas-niveau » permettent de réaliser rapidement des simulations : la réalisation de nombreuses variantes en modifiant, par exemple, les hypothèses de projection démographique et/ou économique ou encore en changeant la réglementation des retraites est ainsi rendue possible dans des délais raisonnables. La précision des résultats obtenus est toutefois plus faible que celle obtenue par des modèles s'appuyant sur de plus gros échantillons (Prisme par exemple).

L'approche par microsimulation permet la modélisation de phénomènes complexes dans lesquels la variabilité des caractéristiques et des comportements individuels joue un rôle important. La microsimulation est dynamique afin d'apprécier l'évolution à long terme des retraites.

2. Sources et architecture du modèle

2.1. Échantillon initial : Patrimoine 2009-2010

Le modèle s'appuie sur un échantillon représentatif de la population française au 1^{er} janvier 2010, construit à partir de l'enquête Patrimoine 2009-2010. Composé d'environ 62 000 individus, cet échantillon initial contient des informations sur les liens familiaux ainsi que l'historique des positions occupées sur le marché du travail. Le tirage de cet échantillon est calé sur des sources extérieures (Enquêtes Emploi, Enquêtes Annuelles de Recensement, Bilan démographique) afin d'être représentatif vis-à-vis des grandeurs suivantes :

- Âge, à un niveau fin ;
- Répartition par niveau de diplôme ;
- Répartition par type de ménage ;
- Taux d'activité par sexe et par tranche d'âge décennale entre 15 et 64 ans ;
- Part de salariés du public, du privé, part de chômeurs et part d'indépendants par sexe et par tranche d'âge (15-24 ans, 25-49 ans, 50-64 ans)

Informations observées:

Sur l'échantillon initial, les informations observées sont : le sexe, l'année de naissance, l'âge de fin d'études (pour les individus ayant terminé leurs études), la catégorie dans la fonction publique (active ou sédentaire) en 2009, le fait d'être né en France ou non.

Informations imputées :

Les informations imputées sont : l'âge de fin d'études (pour les individus non encore sortis des études), la catégorie dans la fonction publique en rétrospectif, le niveau de diplôme. Par ailleurs, certains liens familiaux ont été reconstitués, notamment pour rattacher les enfants hors domicile à leurs parents.

2.2 Une structure sous la forme de trois modules successifs

Le modèle comprend trois modules distincts : le module générateur des biographies démographiques, le module générateur des trajectoires professionnelles et le module « retraite ».

La construction des trajectoires démographiques s'appuie sur les projections démographiques de l'Insee publiées en 2016, couvrant le champ France entière sur la période 2013-2070. Le nombre de décès, de naissances et les flux migratoires sont ainsi calés chaque année de la projection.

Les parcours professionnels des individus sont observés jusqu'en 2009, année de base. À compter de 2010, leurs carrières (statuts d'activité et revenus) sont projetées en respectant des contraintes de calage fondées sur des hypothèses macroéconomiques. Celles-ci portent sur les gains de productivité du travail, le taux de chômage, tous deux repris des publications du COR, et le taux d'activité, issu des projections de population active de l'Insee.

Le module « retraite » simule le départ à la retraite et calcule le montant des droits associés y compris pour les personnes déjà retraitées. Ces droits incluent les droits directs et la pension de réversion des régimes de base (régime général, SSI, fonction publique) et du régime complémentaire Agirc-Arrco. Outre la réversion, la dimension « ménages » permet de simuler l'attribution de l'Allocation de Solidarité aux Personnes Âgées (Aspa).

De nombreuses sources sont mobilisées afin de constituer l'échantillon initial et estimer les équations destinées à alimenter la projection: enquête Patrimoine 2009-2010, enquêtes Emploi, enquêtes annuelles de recensement, bilan démographique, enquête Formation et qualification professionnelle, projections démographiques, enquête Histoire familiale, enquête Génération 1998, EIC, EIR.

3. Description des différents modules

Le module démographie génère les trajectoires démographiques individuelles et les liens familiaux. Les événements simulés sont les unions, les séparations, les naissances, les migrations et les décès.

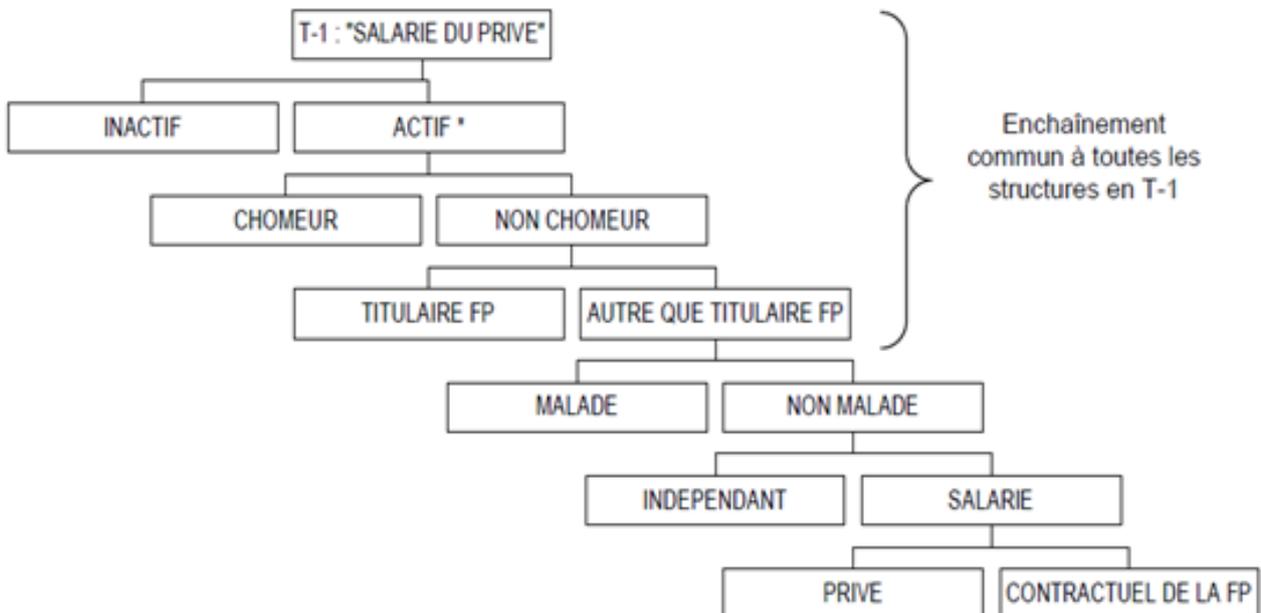
Tous les types d'union sont simulés, sans distinction de leur nature juridique (mariage, PACS ou union libre). Pour les unions, les premières mises en couple et les remises en couple sont modélisées séparément, sur la base de probabilités dépendant de caractéristiques individuelles. L'appariement des conjoints se fait par la minimisation d'une distance portant sur deux dimensions, l'âge et l'âge de fin d'études. Les probabilités de rupture sont calculées uniquement à partir des caractéristiques de la femme (âge à l'union, durée des études) ou du couple (durée de vie en couple, présence d'enfants), mais pas des caractéristiques propres de l'homme. Pour les naissances, l'âge de procréation est compris entre 16 et 45 ans. Les probabilités de naissance sont calculées suivant des équations séparées selon l'âge de fin d'études de la mère, le rang de l'enfant (à partir du 2^e enfant, les équations sont distinctes suivant que l'enfant précédent est issu ou non de l'union en cours). Dans chaque équation, la probabilité dépend essentiellement de l'âge de la mère et de la durée de l'union. La mortalité est simulée à partir d'un quotient de mortalité qui dépend du sexe, de l'âge, de la génération et du niveau de diplôme. La simulation de la migration consiste à ajouter chaque année dans la population des individus supplémentaires avec des caractéristiques démographiques tirées aléatoirement pour reproduire les caractéristiques principales des flux de migrants, en termes de structure par âge et de *sex ratio*. Les entrées concernent les moins de 60 ans hormis les nouveaux-nés et les jeunes adultes, pour qui seules des sorties sont simulées, suivant le même principe de simulation des entrées. Aucune entrée ni aucune sortie n'est simulée pour les plus de 60 ans.

Le module des trajectoires professionnelles simule les transitions sur le marché du travail et les rémunérations. Les statuts modélisés sont : non-cadre du privé, cadre du privé, contractuel de la FP, titulaire FPE en catégorie active, titulaire FPE en catégorie sédentaire, titulaire FPTH en catégorie active, titulaire FPTH en catégorie sédentaire, non-salarié, chômeur, inactif, invalide du secteur privé, invalide du secteur public, invalide non-salarié, en cours de scolarité, AVPF.

Les trajectoires professionnelles sont projetées à partir de 2010 selon un pas annuel. Les probabilités de transition de chaque individu dépendent de ses caractéristiques individuelles observées au moment de la transition. Ces probabilités ont été estimées de manière

séparée sur trois sous-parties de la carrière professionnelle (début, milieu, fin) mais selon une méthodologie commune de logits emboîtés. Concrètement, pour chaque statut observé à l'année t-1, les transitions professionnelles sont estimées par des arbres de décision successifs : le statut en t est celui de sortie de l'arbre.

Illustration 1: Arbre de transition d'une personne salariée du privé en t-1



Outre le calage du taux d'activité et du taux de chômage, les transitions respectent des cibles de part de la FP dans l'emploi total et de part de la FPE dans la FP. Les revenus salariaux sont simulés de manière prospective mais également rétrospective, en l'absence d'informations correspondantes dans l'enquête Patrimoine. La modélisation repose sur une décomposition du logarithme du revenu salarial rapporté au SMPT en une partie déterministe fonction de caractéristiques observables, une composante permanente inobservée et une perturbation idiosyncratique. Les variables explicatives sont essentiellement la durée en emploi, la durée dans le statut courant, l'âge de fin d'études et des indicatrices de transition. Les équations sont distinctes selon le sexe et le secteur (privé, public, non-salariés)

À partir des trajectoires démographiques et professionnelles issues des deux modules précédents, le bloc « retraite » simule le départ à la retraite et le montant des droits associés. Plusieurs options de comportement de liquidation sont proposées par le modèle. Le départ peut avoir lieu à un âge exogène, au taux plein, suivant une cible de taux de remplacement, ou par la maximisation de la somme actualisée des droits à la retraite. Une option de départ selon un modèle de Stock et Wise est aussi implémentée. Sous cette hypothèse, les individus partent à la retraite en minimisant leur *Option Value*, qui est l'écart entre l'utilité d'un départ à la retraite à la date optimale avec l'utilité que l'agent retire d'un départ aujourd'hui. Cette approche permet d'ajouter à la dimension purement financière une dimension associée à l'utilité que procure l'arrêt du travail (ou, de façon symétrique, la désutilité associée au travail).

Les départs à la retraite et les montants de pension sont simulés de manière prospective mais aussi rétrospective, et ce, bien que l'enquête Patrimoine fournisse les âges de départ

et les niveaux de pension associés. Deux raisons expliquent ce choix. La première motivation est de minimiser le risque de discontinuité au démarrage de la projection : d'une part, le modèle de projection ignorant forcément un certain nombre de caractéristiques fines des barèmes, les pensions calculées en flux ne peuvent être totalement cohérentes avec celles qui seraient parfaitement observées en stock ; d'autre part, les pensions observées en stock sont loin de l'être parfaitement en raison de sur- ou sous-déclarations. Une simulation globale homogène du flux et du stock permet une projection à biais constant plutôt qu'à biais variable. Un second intérêt de cette façon de procéder est qu'elle permet la simulation de scénarios fondés sur des conditions initiales contrefactuelles.

4. Méthodes de validation du modèle et contrôles de cohérence

La validation du modèle porte notamment sur les points suivants :

- *Âge de fin études* : comparaison avec les données de l'enquête Emploi pour les générations nées entre 1950 et 1980 ;
- *Répartition des femmes par nombre d'enfants* : comparaison avec l'enquête Patrimoine ;
- *Rupture et remise en couple* : comparaison des probabilités prédites et observées dans l'enquête Histoire familiale ;
- *Naissances*: comparaison des probabilités prédites et observées dans l'enquête Histoire familiale
- *Situations familiales à chaque âge* : comparaison, sur quelques générations, avec les données de l'enquête Histoire familiale
- *Revenus salariaux* : comparaison , sur quelques générations, de la distribution des montants de retraite hors minima de pension avec les données de l'EIR, sur le champ des résidents en France ne percevant pas de pension de régimes de non-salariés.

5. Cas d'usage et développements en cours

5.1 Cas d'usage : Aubert et Bachelet (2012)

Parmi les nombreuses études réalisées à partir du modèle Destinie figure celle de Patrick Aubert et Marion Bachelet en 2012¹. Dans cette étude, les auteurs s'intéressent à la redistribution intragénérationnelle opérée par le système de retraite français, d'une part, en comparant les distributions de pensions et de salaires et en identifiant les bénéficiaires de la redistribution, et, d'autre part, en quantifiant la contribution des divers mécanismes de redistribution, qu'ils soient explicites (comme la MDA ou les minima de pension) ou implicites (comme la règle des 25 meilleures années au régime général ou la limitation à 1 du coefficient de proratisation dans les régimes de base).

Cette étude montre que les mécanismes implicites liés aux règles d'acquisition des droits et de calcul des pensions jouent globalement dans le sens d'un accroissement des inégalités entre retraités d'une même génération, autrement dit que la redistribution opérée par ces mécanismes implicites joue en défaveur des assurés à bas salaires et/ou à carrières courtes (particulièrement pour les femmes). Cependant, les dispositifs de solidarité jouent globalement dans le sens d'une nette réduction des inégalités et, au total, le système de retraite opère une redistribution qui contribue à réduire les inégalités entre retraités.

Ainsi, pour les salariés nés entre 1955 et 1964, alors que le rapport entre le seuil au-dessus duquel se situent les 10 % des personnes ayant perçus sur l'ensemble de leur carrière les salaires les plus élevés et celui en-dessous duquel se situent les 10 % les moins bien payés atteint 5,85, il n'est plus que de 4,10 entre les seuils définissant les 10 % des retraites de droits propres les plus élevées et les 10 % les plus faibles. Les principaux mécanismes qui contribuent à cette réduction des inégalités de carrière, plus marquée parmi les femmes que parmi les hommes, sont la validation de périodes assimilées, la validation de trimestres au titre de l'AVPF et les *minima* de pension. Cette réduction des inégalités est obtenue alors même que les pensions versées par le « cœur du système » sont plus dispersées (rapport inter-décile de 6,66) que les salaires perçus sur l'ensemble de la carrière (rapport inter-décile de 5,85).

¹Aubert, P., & Bachelet, M. (2012). Disparités de montant de pension et redistribution dans le système de retraite français. *Document de travail n° G2012/06, Insee.*

9. Dispersion des montants de pension simulés en prenant en compte successivement les dispositifs explicites et implicites de redistribution

		Écarts entre déciles		
		D9/D1	D5/D1	D9/D5
Dispositifs explicites (durée)	(1) Aucun dispositif : pensions proportionnelles au cumul des salaires	5,85	3,12	1,88
	(2) + mécanismes implicites	6,66	3,41	1,95
	(3) + validation de trimestres au titre des périodes assimilées (chômage, préretraite)	5,65	3,00	1,88
	(4) + validation de trimestres au titre de l'AVPF	5,21	2,78	1,88
	(5) + majoration de durée d'assurance (MDA)	4,90	2,63	1,86
Dispositifs explicites (montant)	(6) + points gratuits (chômage, préretraite et GMP) dans les régimes complémentaires	4,88	2,60	1,88
	(7) + SMIC porté au compte en cas d'AVPF	4,75	2,53	1,88
	(8) + bonification de pension pour 3 enfants et plus	4,71	2,50	1,89
	(9) + minima de pension (minimum contributif et minimum garanti)	4,10	2,18	1,89

Champ : retraités n'ayant pas de pension d'indépendant et ayant travaillé au moins un an, nés entre 1955 et 1964 et résidant en France métropolitaine.

Lecture : par rapport à une situation où les minima de pension ne seraient pas pris en compte, l'application de ces derniers conduit à réduire le rapport interdécile (D9/D1) de 4,71 à 4,10.

Note : les numéros indiqués pour les différents scénarios de simulation correspondent à l'ordre de prise en compte des divers dispositifs dans les simulations successives. Le scénario (1) correspond à une situation où les pensions seraient directement proportionnelles au cumul des salaires sur la carrière.

Source : Insee, modèle *Destinie 2*.

Cette quantification est rendue possible par le recours à la microsimulation : *Destinie* permet en effet de neutraliser tout ou partie des différents mécanismes décrits par les auteurs, ce qu'une approche plus macroéconomique ou en moyenne ne permettrait pas de faire.

5.2 Développements en cours

Destinie est un modèle de microsimulation flexible et évolutif, pouvant connaître des améliorations et des extensions au fil du temps. Le modèle peut être étendu par le développement de modules supplémentaires ou le codage de réglementations nouvelles. *Destinie* peut aussi être rafraîchi grâce à la mise à jour des données sources du modèle. Ces dernières années, un module de santé ainsi qu'un module dépendance ont été intégrés. Parmi les développements à venir du modèle, nous pouvons mentionner les travaux de L. Galiana, T. Guichaoua et L. Wilner portant sur un module d'accumulation du patrimoine financier et le rebasage consistant à intégrer l'enquête Patrimoine 2017-2018 en lieu et place de Patrimoine 2009-2010 en tant qu'échantillon de base. Enfin, des réflexions liées à la réforme des retraites sont aussi en cours.

Illustration 2: Schéma récapitulatif de la structure de Destinie

