

# Introduction

## Microsimulations statique et dynamique appliquées aux politiques fiscales et sociales : modèles et méthodes

Didier Blanchet \*, Cyrille Hagneré \*\*, François Legendre \*\*\*  
et Florence Thibault \*\*\*\*

L'Acoss, la Cnaf, l'Erudite de l'Université Paris-Est et l'Insee ont coorganisé en 2013 une journée de travail visant à faire le point sur l'évolution des outils de microsimulation développés et utilisés aussi bien au sein de l'administration française que dans des cadres plus académiques. Cette journée reprenait le principe d'une journée de même type qui avait donné lieu en 2003 à un numéro spécial de la revue *Économie et prévision* (Legendre, Lorgnet et Thibault, 2003). La journée qui s'est déroulée au printemps 2013 donne cette fois lieu à une double valorisation, partagée entre la présente livraison d'*Économie et statistique* et un numéro à paraître de la *Revue économique*. Cette double valorisation combine une sélection de textes issus de ce colloque et de plusieurs autres travaux ayant débouché postérieurement. Le numéro spécial de la *Revue économique* présentera plusieurs applications de la microsimulation à l'évaluation de politiques publiques. En matière d'évaluation des politiques publiques, les angles actuellement dominants sont soit celui de l'expérimentation sur le modèle des essais cliniques, avec un groupe témoin et un groupe « traité », soit celui de l'évaluation microéconométrique *ex post* de politiques en place, à partir de situations d'expériences naturelles. Or l'expérimentation ne répond pas à tous les besoins d'évaluation en vraie grandeur et les méthodes *ex post* ne peuvent répondre qu'indirectement et partiellement aux nombreux besoins d'évaluation *ex ante*. Ces besoins d'évaluation *ex ante* ne sont pas davantage couverts par les modèles macroéconométriques car les politiques à évaluer, même lorsqu'elles sont générales, s'adressent *in fine* à des publics particuliers. Les décideurs ont besoin d'études d'impact préalables qui, *a minima*, dénombrent le public éligible au nouveau dispositif et chiffrent son coût, ce qui nécessite des modèles fondés sur des données microéconomiques. C'est ce que font les modèles de microsimulation et ils ont donc un rôle important à jouer dans ces travaux d'évaluation *ex ante*, avec l'atout de s'intéresser par nature non seulement aux effets moyens des politiques mais aussi à la dispersion de leurs conséquences.

### 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.

\* Insee.

\*\* Agence centrale des organismes de sécurité sociale (Acoss).

\*\*\* Erudite, Université Paris-Est et Acoss.

\*\*\*\* Cnaf, Département des statistiques, des prévisions et des analyses.

L'application à l'évaluation des politiques publiques sera également très présente dans ce numéro d'*Économie et Statistique*, mais ce sera davantage en toile de fond car l'accent sera davantage méthodologique. La microsimulation est une méthode relativement simple dans son principe, mais les microsimulateurs n'en sont pas moins confrontés à divers problèmes techniques auxquels les réponses restent peu mutualisées car souvent peu documentées dans les articles présentant les résultats de ces modèles, et avec peu d'ouvrages de synthèse auxquels se référer. Par exemple, ce n'est que depuis peu de temps qu'on dispose d'un *handbook* proposant un tour d'horizon systématique des différents aspects de la méthode (O'Donoghue, 2014). On peut voir le présent numéro comme contribuant lui aussi à combler ce besoin de documentation méthodologique, à sa manière, en proposant d'ouvrir davantage les boîtes noires d'un certain nombre de ces modèles, existants ou encore en cours de développement. Les champs qui seront couverts sont les champs d'application les plus classiques de ce type de modélisation : fiscalité et redistribution, retraite, dépendance, santé... Selon une typologie usuelle, les modèles présentés seront soit statiques soit dynamiques. Les modèles statiques cherchent à reproduire l'impact des politiques sur les situations individuelles à date donnée. Les modèles dynamiques se proposent de retracer les trajectoires individuelles en engendrant artificiellement les principaux événements du cycle de vie, tels que la fin des études, la mise en couple, le passage éventuel par le chômage... Ces modèles dynamiques peuvent se subdiviser en deux sous-catégories, celle des modèles « de période » qui projettent l'ensemble d'une population au cours du temps et celle des modèles « de cohorte » qui projettent une génération sur l'ensemble de son cycle de vie.

### Un échantillon de sept modèles

Commençons par une présentation rapide des différents articles qui composent ce numéro. Dans la première catégorie des modèles statiques, **Antoine Bozio**, **Malka Guillot** et **Quentin Laffeter** proposent tout d'abord une présentation du modèle Taxipp, développé à l'Institut des politiques publiques (IPP), où il est à la fois un lointain descendant du modèle pionnier Sysiff (Bourguignon *et al.*, 1988) et l'héritier direct du calculateur fiscal dont la mise en ligne avait accompagné la sortie de l'ouvrage *Pour une révolution fiscale* publié en 2011 par Thomas Piketty, Camille Landais et Emmanuel Saez. L'étude de la fiscalité et des réformes fiscales est de fait un champ d'application très usuel de ces modèles statiques. Le deuxième article du numéro applique cette même approche statique à un domaine plus inhabituel puisqu'il s'agit d'un sujet généralement envisagé à long terme, celui de la dépendance, qu'on s'attend donc à voir traiter à l'aide de modèles dynamiques. Mais c'est oublier que la dépendance est aussi un problème actuel, avec un mode de prise en charge qui peut continuer à évoluer et qui peut donc aussi faire l'objet de réformes de court terme à effets immédiats. C'est l'objectif du modèle Autonomix de la Drees de permettre leur simulation et il est présenté dans l'article d'**Arnaud Fizzala**.

Tout le reste du numéro est consacré à des travaux de microsimulation dynamique, sous différentes formes. Il enchaîne tout d'abord avec trois exemples du domaine où la méthode s'est le plus naturellement imposée, celui de la retraite. Le premier article est celui d'**Yves Dubois** et **Anthony Marino** qui utilise le modèle Destinie pour des évaluations du taux de rendement interne du système de retraite pour les assurés, dans une perspective à la fois inter et intra-générationnelle. Le deuxième article est celui de **Caroline Berteau-Rapin**, **Paul Beurnier** et **Émilie Deynarolles** qui présentent de

nouveaux développements d'un autre modèle lui aussi bien installé, le modèle Prisme construit à la Caisse nationale d'assurance vieillesse pour projeter les droits de l'ensemble de ses affiliés, aussi bien pour des besoins de gestion à court terme que pour les exercices de projections de long terme. Le troisième article présente un nouveau venu dans ce paysage des modèles de retraite, le modèle Trajectoire de la Drees, présenté par **Cindy Duc, Félix Housset, Laurent Lequien et Corentin Plouhinec**, qui est un modèle tous régimes basé sur l'*Échantillon interrégimes de cotisants (EIC)* géré par le même organisme.

Les trois derniers articles appliquent la microsimulation dynamique à d'autres champs que la retraite. L'article de **Pierre Courtioux et Vincent Lignon** s'appuie sur un modèle de cohorte, le modèle Gameo de l'Edhec, dont la vocation de départ est l'étude des effets redistributifs sur cycle de vie de différents modes de financement de l'enseignement supérieur (Courtioux *et al.*, 2011), un type d'application qui sera aussi illustré dans le numéro à paraître de la *Revue économique* (Allègre et Timbeau, 2016). Ce modèle Gameo est ici étendu et appliqué à une autre problématique : celle des effets de l'homogamie sur la répartition des revenus des ménages. **Claire Marbot et Delphine Roy** reviennent ensuite sur le même sujet que celui traité par Fizzala, celui de la dépendance, mais cette fois sous l'angle dynamique de la projection à long terme, à l'aide d'un module *ad hoc* ajouté au modèle de retraites Destinie. Ce module leur permet de présenter des résultats qui concernent à la fois la situation familiale des personnes dépendantes et le coût financier de l'Allocation personnalisée d'autonomie (APA), sous différents scénarios d'indexation de son barème. Le tout dernier texte du numéro concerne la santé. Il présente les premières étapes de la construction d'un module de projection de l'état de santé et des dépenses de santé – ici les soins de ville, ce qui pose des problèmes de modélisation spécifiques : choix d'un indicateur synthétique de santé individuelle, et modélisation de la dépense conditionnelle à ce niveau de santé, dépense qui présente la double caractéristique d'être parfois nulle ou à l'inverse très élevée. Ce dernier article est celui de **Charlotte Geay, Grégoire de Lagasnerie et Makram Larguem**.

Les huit articles couvrent ainsi sept modèles différents. Ils ne représentent qu'une partie de l'existant dont le tableau ci-après essaye de donner une vision un peu plus complète quoique sans garantie d'exhaustivité. Certains des autres modèles mentionnés dans ce tableau seront évoqués dans cette introduction qui ne va pas reprendre séquentiellement les enseignements des huit articles qui vont suivre. Elle va plutôt en resituer l'apport dans le cadre d'un tour d'horizon méthodologique général de la méthode et des problèmes qu'elle soulève. On comprendra mieux ainsi, au fil de la présentation, à la fois le « pourquoi » des questions posées dans les différents articles et le « comment » des solutions apportées par les différentes équipes de modélisateurs. Ceci permettra également de donner un aperçu des travaux qui seront publiés dans le numéro joint de la *Revue économique*.

## **Modèles statiques**

Commençons par un bref retour en arrière, à l'approche du sixantième anniversaire de ce qui est en général présenté comme la contribution fondatrice de cette littérature, l'article publié par Guy Orcutt en 1957 dans *Review of Economics and Statistics*, qui a été le premier à proposer de rendre compte de la dynamique de systèmes

socio-économiques complexes par la simulation informatique de leurs composants élémentaires. Orcutt opposait sa démarche à celle de l'individu représentatif adoptée par la modélisation macroéconométrique et également aux approches de type sectoriel à la Léontief, qui étaient l'une et l'autre en plein développement. L'idée d'Orcutt était de rendre compte des interactions entre les comportements d'agents individuels ou collectifs correspondant aux principaux types d'acteurs économiques : ménages, entreprises, acteurs publics. L'approche devait être dynamique et itérative : à chaque instant, les agents adoptent des comportements ou prennent des décisions sur la base de l'état du monde qu'ils constatent, et il en résulte un nouvel état du monde pour la période suivante.

Le projet initial d'Orcutt était très ambitieux compte tenu des moyens informatiques de l'époque. Le recours à la microsimulation dynamique s'est assez naturellement développé dans le domaine démographique, notamment pour la simulation du processus de

Tableau  
**Modèles de microsimulation appliqués à la fiscalité ou aux politiques sociales**

	Organisme	Domaine d'application
Modèle statiques appliqués aux transferts sociaux		
Ines**	Insee et Drees	Fiscalité et transferts sociaux
Myriade	Cnaf	Fiscalité et transferts sociaux
Saphir	DG Trésor	Fiscalité et transferts sociaux
Euromod	Université d'Essex et Idep (Univ. Aix-Marseille) pour le volet français	Fiscalité et transferts sociaux
Taxipp*	Institut des politiques publiques	Fiscalité et transferts sociaux
Openfisca	Etalab, France-Stratégie, Idep et IPP	Fiscalité et transferts sociaux
Isis	DG Trésor	Dépenses de santé
Omar	Drees	Dépenses de santé
Autonomix*	Drees	Dépendance
Modèles dynamiques de période appliqués principalement à la retraite		
Destinie*	Insee	Retraites tous régimes, dépendance et santé (en coll. avec la DG Trésor)
Prisme*	Cnav	Retraites du régime général
Trajectoire*	Drees	Retraites tous régimes
Pensipp**	Institut des politiques publiques	Retraites tous régimes
Aphrodite	DG Trésor	Retraites tous régimes
Pablo	Service des retraites de l'État	Retraites des agents titulaires de la fonction publique
Modèles dynamiques de cohorte		
Gameo*	Edhec	Education, fiscalité
Taxipp-life	Institut des politiques publiques	Fiscalité
Modèles statiques ou dynamiques appliqués au marché du travail		
Iris	DG Trésor	Assurance chômage
Matis	DG Trésor	Politiques d'emploi
Ambre**	Acoss	Politiques d'emploi
Worksim**	-	Politiques d'emploi

Note : ce tableau se limite aux modèles pérennes, en général portés par une administration ou un centre de recherche (à l'exception du modèle Worksim). De nombreux modèles sont également développés de manière indépendante à l'occasion de travaux d'étude spécifiques.

\* Modèles représentés dans ce numéro

\*\* Modèles représentés dans le numéro joint de la *Revue Économique*.

formation des familles<sup>1</sup>. Dans le domaine économique, en revanche, à quelques exceptions près (Bennet et Bergmann, 1974 ; Eliasson, 1977) beaucoup d'applications opérationnelles se sont longtemps cantonnées à l'approche statique et comptable, avec des modèles consacrés à l'analyse de la fiscalité et des transferts sociaux. Pour la France, le premier en date a été le modèle Sysiff (Bourguignon *et al.*, 1988) qu'on a déjà présenté plus haut comme l'ancêtre du modèle Taxipp présenté ici par Bozio et ses co-auteurs. Plusieurs autres modèles de ce type étaient mobilisés dans le numéro spécial de 2003 de la revue *Économie et prévision* (Legendre *et al.*, 2003 ; Bourguignon et Spadaro, 2003) dont le modèle Ines, développé et utilisé par l'Insee et la Drees (Albouy *et al.*, 2003), et le modèle Myriade de la Cnaf (Legendre *et al.*, 2001). Le numéro précédent d'*Économie et Statistique* a d'ailleurs présenté un travail issu de Myriade, consacré au taux de couverture du coût de l'enfant selon deux définitions de ce dernier (Favrat, Marc et Pucci, 2015). Un travail basé sur le modèle Ines figurera de son côté dans le numéro associé de la *Revue économique* : il sera consacré à la prise en compte de l'« avantage HLM » dans la mesure de la redistribution (Cazenave, Domingues et Vicard, 2016). Le numéro à venir de la *Revue économique* présentera aussi un autre exemple d'application de la microsimulation à la fiscalité, avec une étude consacrée aux effets des réformes de la décennie 2000 sur les droits de succession (Goupille, 2016). On citera encore le modèle Saphir de la DG Trésor dont on trouvera là aussi une application dans un numéro récent d'*Économie et Statistique*, consacrée à l'évaluation du niveau de vie des personnes rémunérées au Smic (Favrat, Prady et Tavan, 2011)<sup>2</sup>. On citera enfin le volet français du modèle européen Euromod (Sutherland et Figari, 2013), porté par l'Idep (Université d'Aix-Marseille) et le modèle OpenFisca développé par Etalab et France-Stratégie, avec des appuis de l'Idep et de l'IPP. À ce stade, ce dernier modèle est davantage un calculateur fiscal qu'un modèle de microsimulation à proprement parler car il n'est pas attaché à une base de données représentative de la population française, mais il peut le devenir facilement par raccordement à un tel fichier.

La modélisation statique est aussi utilisée pour l'évaluation de la prise en charge des soins de santé. Pour un *survey* de ce type d'application et un exemple remontant aux années 1990 on pourra se référer à Breuil-Genier (1998) et Lachaud *et al.* (1998). Pour des exemples français plus récents, on se référera à Debrand et Sorasith (2010), Lardellier *et al.*, (2012) et Geoffard et de Lagasnerie (2012). Ce dernier article avait été consacré à la simulation de réformes de grande ampleur du système de remboursement des soins : le numéro à venir de la *Revue économique* présentera un prolongement de ce modèle et de ces simulations, toujours statique, mais avec endogénéisation des comportements de consommation de soins (de Lagasnerie, 2016).

Le fait que la microsimulation se soit d'abord imposée pour ce type d'application est facile à comprendre. Les barèmes fiscaux et sociaux sont des barèmes complexes et les effets qu'on veut mesurer sont des effets au niveau des individus ou des ménages. La démarche la plus naturelle pour traiter ces problèmes est donc bien de partir de fichiers de ménages, avec les diverses caractéristiques en termes de revenu et d'emploi de leurs membres puis à simuler, sur ces ménages, l'impact de règles variables de prélèvements et de transferts. On en dérive les gains ou pertes individuels à la modification de ces transferts, ainsi que leur bilan total pour les finances publiques. Cela a un sens de le faire

---

1. Pour des vues d'ensemble sur ces applications démographiques, le lecteur pourra par exemple se référer à Van Imhoff et Post (1997) ou Morand *et al.* (2010).

2. Pour d'autres applications du même modèle on pourra aussi consulter Chanchole et Lalanne (2012) ou Boisnault et Fichen (2015).

dans un cadre statique donnant une photographie de ces transferts à une date donnée et permettant sa comparaison avec les photographies qui résulteraient de la mise en œuvre d'autres barèmes fiscaux et sociaux. La microsimulation remplace ainsi une collection de cas-types dont il est toujours difficile de savoir si elle couvre bien toutes les situations et ne focalise pas à tort sur des populations peu représentatives. Comme on l'a indiqué, l'article de Fizzala montre que la démarche est tout aussi légitime pour un sujet tel que la dépendance, pour lequel on peut aussi se poser des problèmes d'évaluation de politiques à effets potentiellement immédiats.

### **Introduire une dimension temporelle : les modèles dynamiques**

Le passage à l'approche dynamique consiste à ajouter à ces modèles une dimension temporelle. La dynamisation peut répondre à deux besoins et se faire de deux manières, détaillées par exemple par Dekkers (2015).

Un premier besoin est l'actualisation des résultats d'un modèle s'appuyant sur une base de données plus ancienne que la période pour laquelle on souhaite avoir des résultats. C'est ce que décrivent Bozio et ses co-auteurs pour le modèle Taxipp, ce type de dynamisation étant également appliqué de manière régulière à des modèles statiques tels qu'Ines ou Myriade. Le modèle Ines vient par exemple d'être aménagé pour produire une estimation avancée des principaux indicateurs d'inégalité de niveau de vie, notamment le taux de pauvreté. Cette estimation sera publiée un an après la fin de l'année considérée, avec une première estimation provisoire et expérimentale publiée en décembre 2015 pour les données de l'année 2014.

Pour ce genre de projection à court-terme, une technique possible consiste à simplement repondérer les individus de la base initiale sans changements de leurs caractéristiques. On peut préférer parler dans ce cas là de « pseudo-dynamisation ». Le terme de microsimulation dynamique *stricto sensu* s'applique plutôt à la deuxième façon de faire évoluer les données du modèle dans le temps, lorsque ce sont les caractéristiques des différents individus simulés qu'on modifie à chaque étape de la simulation, d'une façon qui cherche à reproduire les différents événements ou changements d'états auxquels ces individus sont exposés. Cette forme de dynamisation s'impose tout particulièrement lorsqu'on a besoin de simuler des transferts dont l'incidence ne s'apprécie que dans la durée, ce qui sera typiquement le cas des retraites.

Dans sa première version, celle des modèles « de période », cette dynamisation consiste à faire vieillir d'une année sur l'autre les individus de l'échantillon de départ par combinaison de règles déterministes (telles que l'incrémentation de l'âge), de transitions pseudo-aléatoires entre les différents états que simule le modèle (par exemple une transition de l'emploi vers le chômage ou l'inactivité) ou éventuellement de comportements plus complexes (par exemple le choix d'un âge de départ en retraite compte tenu des droits qu'offre le système aux différents âges de départ possibles). Une bonne partie du contenu des articles de ce numéro est consacrée à détailler la forme que donne chaque modèle à ces différentes règles. Évidemment, se borner à ce vieillissement conduit à simuler un échantillon qui se vide progressivement de ses tranches d'âge les plus jeunes – ce qui est la limite du modèle présenté dans l'article de Geay et ses co-auteurs. Il convient donc d'y rajouter la création de nouveaux individus, qu'on peut soit introduire *ex nihilo*, ce que peut faire un modèle

ne s'intéressant qu'aux seuls individus indépendamment de leurs liens familiaux, soit en tant qu'enfants des couples appartenant à l'échantillon simulé, comme le fait notamment le modèle Destinie.

Dit d'une autre façon, le principe de ces modèles de période est donc de partir d'un échantillon représentatif à une date  $t_0$  et de le faire vieillir et de le renouveler de proche en proche de la date  $t_0+1$  à un horizon  $t_0+h$ . Si l'échantillon de départ est un fichier d'enquête, ou tout du moins une sélection de variables d'un tel fichier, cela veut dire qu'on obtient en sortie des fichiers équivalents à ce fichier d'enquête pour toutes les dates futures, qui sont donc susceptibles des mêmes exploitations statistiques, avec une variété d'outputs potentiels bien plus élevée que pour n'importe quelle autre forme de projection.

Il existe de très nombreux modèles de ce type à l'étranger (voir le *survey* récent de Li et O'Donoghue, 2013). Dans le cas de la France, les principaux modèles sont le modèle Destinie de l'Insee déjà cité (voir notamment Bonnet et Mahieu, 2000 ; Bardaji *et al.*, 2003 ; Blanchet *et al.*, 2011), le modèle Prisme de la Cnav mais aussi désormais le modèle Trajectoire de la Drees, tous trois représentés dans ce numéro. Mais la liste ne s'arrête pas à ces trois modèles. Un autre modèle généraliste cousin à la fois de Taxipp et du modèle Destinie de l'Insee est le modèle Pensipp développé comme Taxipp à l'Institut des politiques publiques et qui sera appliqué, dans le numéro associé de la *Revue économique*, à la simulation de différents scénarios de réforme structurelle ou paramétrique du système de retraite français (Blanchet, Bozio et Rabaté, 2016). Il existe aussi un modèle de projection des droits à retraite en cours de finalisation à la DG Trésor, le modèle Aphrodite. Et la plupart des régimes importants commencent aussi à s'équiper de modèles de ce type pour leurs propres besoins : le modèle Pablo est en cours de construction au Service des retraites de l'État, et la construction de modèles de microsimulation est aussi à l'étude à la CNRACL et pour les régimes complémentaires Arrco/Agirc.

Même si la question des retraites est prédominante parmi les usages des modèles dynamiques, ils sont par ailleurs mobilisés sur l'ensemble des questions liées au vieillissement démographique telles que la prospective des dépenses de santé. Pour un point des travaux qui existaient au début des années 1990, on pourra à nouveau se référer à Breuil-Genier (1998). Des exemples d'applications françaises plus récentes seront trouvés dans Dormont, Grignon et Huber (2006) ou Thiébault, Barnay et Ventelou (2013). Un sujet associé est celui de la dépendance : il avait été abordé une première fois à l'aide du modèle Destinie au milieu des années 2000 (Duée et Rebillard, 2004) : l'article de Marbot et Roy qui est présenté dans ce numéro est un prolongement de ce premier travail.

Mais cette façon de dynamiser la microsimulation n'est pas la seule. La seconde est celle des modèles « de cohorte » qui se focalisent sur une ou des génération(s) particulière(s), ce qui est la meilleure façon de faire lorsqu'on s'intéresse à la dynamique des inégalités ou de la redistribution au niveau intragénérationnel. Il est certes possible de suivre des générations à l'intérieur d'un modèle de période, c'est ce que font ici Dubois et Marino pour évaluer le retour sur contributions qui est assuré par le système de retraite français. Mais ils tirent pour cela parti du fait que Destinie inclut aussi une simulation rétrospective des carrières passées de l'ensemble des individus présents l'année de base et peut recalculer les retraites des individus ayant déjà liquidé sous des hypothèses

contrefactuelles de législation. Un modèle de période pur ne simulant que d'une année de base  $t_0$  à une année horizon  $t_0+h$  ne simule aucune génération complète si  $h$  n'est que de quelques dizaines d'années. Un modèle de cohorte renonce à l'objectif de suivre parallèlement l'ensemble des générations coexistant à la date  $t_0$ , il n'en retient qu'une seule mais, en revanche, il la suit jusqu'à son décès. Comme exemple, outre le modèle Gameo qui sera présenté plus loin dans l'article de Courtioux et Lignon, on peut notamment citer le modèle *Taxipp-life* lui aussi en cours de développement à l'IPP qui est une extension sur cycle de vie du modèle Taxipp.

Les modèles de période ou de cohorte peuvent différer sur d'autres aspects. D'une part, raisonner en termes de cohortes permet éventuellement d'appliquer ce que les démographes qualifient de méthode de la cohorte fictive, c'est-à-dire imaginer ce que serait le destin d'une cohorte soumise toute sa vie durant aux conditions socio-économiques et institutionnelles d'une période de référence donnée, ce qui peut avoir un intérêt intrinsèque. Ce n'est toutefois pas ce qui sera fait ici dans l'article de Courtioux et Lignon, qui se centrent sur les individus nés en 1970 mais leur simulent de vraies trajectoires prospectives tenant compte d'hypothèses complètes d'évolutions temporelles des salaires ou de l'état du marché du travail, exactement comme le font les modèles de période.

Une autre différence qui recoupe l'opposition période/cohorte est la distinction entre approches en populations « fermées » ou « ouvertes ». Ici, les termes « fermée » et « ouverte » n'opposent pas les modèles qui prennent ou ne prennent pas en compte les flux migratoires, ils opposent deux façons de formaliser la mise en couple et les autres liens familiaux. Cette mise en couple joue un rôle dans le texte de Dubois et Marino, puisqu'ils simulent *in fine* des taux de rendement du système de retraite incluant les pensions de réversion, et elle est centrale dans l'article de Courtioux et Lignon puisque l'étude des effets de la mise en couple et de l'homogamie est l'objet même de leur étude. L'approche en population fermée qui est celle de Destinie signifie que les mises en couple sont simulées entre individus qui sont membres de la population d'intérêt, en appariant des individus de cette population selon une modélisation – sommaire – d'un processus matrimonial. Ce sont ensuite les enfants issus de ces couples qui renouvellent la population à sa base, avec le complément qu'apporte également la simulation de migrants. Le devenir de ces individus est ensuite simulé, à part entière, en combinant règles d'évolution individuelles pour ce qui concerne leurs histoires propres, et règles conjointes pour les événements dépendant des autres personnes auxquels ils sont liés. C'est de cette façon qu'on construit des structures familiales complètes qui sont utilisées par Marbot et Roy pour projeter l'entourage familial des personnes dépendantes.

Ce genre d'approche n'est pas transposable à un modèle de cohorte puisqu'on n'y modélise qu'au plus quelques générations adjacentes : les individus qui sont liés aux individus de la cohorte d'intérêt ne sont pas eux-mêmes membres de la cohorte d'intérêt, ou tout du moins pas nécessairement. Il est donc plus pratique de les gérer comme individus « additionnels », qui entrent et sortent du champ de la simulation au gré des besoins – c'est le sens du mot « ouvert » – et sans qu'il soit nécessaire de les simuler aussi complètement que les individus qui sont l'objet de l'étude. Les seules caractéristiques de ces individus qui nous intéressent sont celles dont on a besoin pour bien décrire les conditions de vie des individus observés. C'est de cette façon que Courtioux et Lignon gèrent les parcours matrimoniaux des individus de leur génération 1970.

## **Microsimulation versus approches « cellulaires » ou « matricielles » : avantages et inconvénients**

Comme on le voit la méthode s'est donc largement diffusée, sous ses différentes formes. Faut-il en conclure que la microsimulation est devenue incontournable quel que soit l'aspect des politiques sociales ou fiscales auquel on s'intéresse ? La question mérite qu'on s'y attarde car elle reste en partie débattue. Elle est certes facile à trancher dans le cas des modèles statiques appliqués à la fiscalité ou aux transferts. Il s'agit de domaines où l'approche par agent représentatif n'a par nature pas de sens : ceci explique que le recours à la simulation y ait été précoce.

Le cas où la question reste en partie ouverte est celui de la microsimulation dynamique. La raison est que la méthode est coûteuse et ceci peut peser en faveur de modèles de granularité intermédiaire entre l'approche totalement agrégée des modèles macro et l'approche totalement individuelle de la microsimulation. Ces modèles qu'on peut qualifier de « cellulaires » ou « matriciels » regroupent les individus en catégories ou cellules plus ou moins fines et raisonnent de manière agrégée à l'intérieur de chacune de ces cellules.

L'exemple de base de cette approche cellulaire est la projection démographique qui consiste à faire évoluer d'année en année un vecteur donnant la répartition de la population par sexe et âge, sur la base de paramètres qui sont les taux de mortalité et de fécondité par sexe et âge, et les flux migratoires eux aussi ventilés par sexe et âge. Les démographes parlent de méthode « des composants ». On peut parler de méthode matricielle parce que la simulation de l'évolution de la population se ramène à appliquer chaque année une matrice de passage transformant le vecteur qui donne l'effectif par sexe et âge de l'année  $t$  au vecteur qui donne celui de l'année  $t+1$ . On peut en dériver des projections appliquées à divers domaines en y plaquant des prévisions de taux ou niveaux moyens pour différents phénomènes d'intérêt au sein de chaque cellule par sexe, âge et période. Les démographes parlent dans ce cas de « perspectives dérivées ». Les modèles de retraite qui ont été utilisés jusqu'au début des années 1990 étaient de ce type, principalement la maquette Margaret de la Direction de la prévision (Vernière, 1990) dans laquelle la population n'était ventilée que selon le sexe et l'âge, l'individu représentatif de chacune de ces cellules étant supposé avoir son taux d'activité moyen et son salaire moyen, sur la base desquels étaient calculées des droits à retraite dans chaque génération, tout cela pour une version très stylisée du système français, supposé homogène et constitué d'un seul pilier prenant en compte l'ensemble du salaire.

Ce type de modélisation reste pertinent : il peut par exemple servir à simuler rapidement l'incidence de certaines des règles du système actuel ou de variantes possibles de ces règles, telles que les règles d'indexation. Et il peut continuer à être envisagé pour la simulation de règles plus complexes, moyennant démultiplication des catégories utilisées pour caractériser les cellules. C'est la voie qui a été suivie récemment par le modèle Promess de la Drees (Aubert *et al.*, 2010) utilisant un découpage très fin de la population selon les durées cotisées dans les différents types de régimes.

Cette approche peut résoudre de nombreux problèmes : le modèle Promess a notamment été utilisé pour simuler l'impact sur les âges de liquidation de réformes des conditions d'accès à la retraite à taux plein. Mais cette façon de procéder finit tout de même par se heurter à une barrière. Les barèmes de calcul des droits à retraite sont en effet

extrêmement complexes et il est impossible de totalement rendre compte de leurs effets en se contentant de résumer les caractéristiques individuelles par le croisement d'un nombre fini de catégories. L'exemple-type est le passage à un calcul de la retraite sur la base des 25 plutôt que des 10 meilleures années de la carrière, qui a été l'une des mesures de la réforme de 1993 : l'impact d'une telle mesure est très variable selon le profil de carrière et son effet moyen dépend donc de la distribution de ces profils de carrière. On peut certes simuler l'impact de ce genre de mesure de manière approchée en construisant une typologie des profils de carrière et en faisant évoluer au cours du temps leurs poids relatifs mais, quand on en arrive à ce degré de complexité, autant basculer dans une approche totalement individuelle, ce qui est le principe de la microsimulation dynamique. Cela avait été la motivation initiale de la construction de Destinie. Berteau-Rapin et ses co-auteurs utilisent le même argument de la complexité des règles pour expliquer que Prisme ait choisi d'entrée de jeu la voie de la microsimulation. Symptomatiquement, la même voie a finalement été retenue par la Drees avec la construction du modèle Trajectoire présenté dans ce numéro, qui permet de donner une réponse aux questions que le modèle Promess ne peut pas traiter.

Au-delà de cet argument, on peut souligner un autre avantage de la microsimulation que viennent illustrer les deux derniers articles du numéro, celui de la plus grande facilité qu'il y a à enrichir ces modèles pour la simulation de caractéristiques non prévues au départ. Comment ce genre d'enrichissement serait-il géré dans le cas d'un modèle matriciel ? Rajouter un critère tel que la dépendance ou la santé à croiser avec l'ensemble des critères déjà gérés par le modèle se fait en reconfigurant les vecteurs ou matrices décrivant l'état de la population et les transitions entre états successifs, ce qui peut-être très coûteux. En revanche, dans un modèle de microsimulation dont le principe est de gérer des enregistrements de caractéristiques individuelles, il est relativement facile d'ajouter une caractéristique nouvelle pour chaque individu et de la faire évoluer en fonction des caractéristiques déjà simulées par le modèle, sans avoir à revenir sur le reste de la structure de ce modèle, et en bénéficiant automatiquement des croisements entre cette ou ces nouvelle(s) variable(s) et les variables déjà présentes.

C'est typiquement le principe suivi dans ce numéro par Marbot et Roy : dans le modèle Destinie initialement appliqué aux seules retraites, il a été possible de rajouter un module simulant l'entrée dans la dépendance, conditionnée par des variables telles que le sexe, l'âge et le niveau d'études, puis de bénéficier du croisement entre cette nouvelle variable et le calcul des droits à retraite pour simuler l'impact d'une prestation dépendance soumise à conditions de ressources. C'est selon le même principe qu'a pu dernièrement être entreprise, sur le même modèle, une greffe des modules de simulation de l'état de santé et du niveau des dépenses présentés ici par Geay et ses co-auteurs (Geay, Koubi et de Lagasnerie, 2015).

Dans la pratique, ce type d'opération s'avère évidemment un peu plus lourd que notre présentation un peu optimiste ne le suggère, car il faudra prendre en compte des phénomènes de rétroaction sur la projection de départ : par exemple simuler la dépendance ou l'état de santé conduit à postuler dans les deux cas une mortalité différentielle selon la valeur de ces deux variables et la combinaison de ces deux mortalités ne redonnera pas nécessairement la mortalité générale simulée dans le modèle initial, ce qui impose des recalages sur la méthodologie desquels on reviendra dans un instant. De plus, la programmation d'une prestation sous conditions de ressources peut s'avérer assez complexe et révéler des fragilités du modèle de départ quant à la projection des inégalités de

revenu. Mais ce type d'ajout est de toute manière quasiment impossible à traiter par des approches agrégées et rend donc la microsimulation quasiment indispensable.

De manière plus générale, on peut dire que c'est un avantage de la méthode que de pouvoir récupérer et combiner facilement des modules de simulation construits indépendamment les uns des autres. Ceci est illustré également par la contribution de Duc *et al.* puisque Trajectoire intègre deux modules préexistants. Le premier est le module de simulation des comportements de départ en retraite qui était déjà celui de Promess : ce dernier calculait une distribution des âges de départ au sein de chacune de ses cellules élémentaires, Trajectoire fonctionne en tirant des âges individuels au sein de ces distributions. L'autre est le programme Caliper (Lermechin *et al.*, 2011 ; Duc *et al.*, 2013), également de la Drees, qui calcule des droits à retraite d'après barème pour tout type de carrière. Trajectoire utilise ce module pour calculer les pensions de droits directs à âge de liquidation connu. Ceci permet au passage de signaler une autre complémentarité entre outils de modélisation, celle qui existe entre microsimulation et outils de simulation de cas-types, déjà évoqués plus haut. Un modèle de microsimulation gagne à s'appuyer sur un outil de calcul de cas-types ayant pu être testé sur des exemples bien contrôlés. À l'inverse, les modules de calcul de droits d'un modèle de microsimulation quelconque peuvent facilement être sortis de ce modèle pour être appliqués à des micro-populations fictives de quelques cas-types dont on postule les profils de carrière et autres caractéristiques.

Tous ces avantages en termes de flexibilité sont toutefois contrebalancés par un inconvénient majeur auquel échappe l'approche matricielle et qui est le caractère stochastique du résultat obtenu : ce que produit une microsimulation n'est qu'une réalisation particulière du processus aléatoire gouvernant l'évolution de la population considérée, sur un échantillon de taille finie. Si le taux de survie d'une sous-population comportant 100 individus est de 85 %, cette sous-population comportera 85 individus avec la méthode matricielle. Avec une microsimulation, cette sous-population pourra compter 84, 85, 86 individus voire un nombre plus éloigné encore de 85 du fait de l'aléa inhérent à la microsimulation. Un arbitrage est donc nécessaire entre l'avantage de pouvoir simuler de manière souple des phénomènes complexes et l'inconvénient de ne le faire qu'avec des résultats bruités.

De quel côté doit pencher cet arbitrage ? Si l'objectif est la projection d'une structure simple telle que la structure par sexe et âge sur la base d'hypothèses de mortalité et de fécondité qui ne dépendent que de l'âge, il n'y a pas de débat sur la bonne façon de faire : la projection matricielle reste préférable, la reproduire par microsimulation ne ferait que redonner le même résultat sous forme bruitée. Dans le cas extrême inverse de projections très complexes, la microsimulation est irremplaçable. Dans tous les cas intermédiaires, le choix est à apprécier au cas par cas. Cet arbitrage peut cependant tenir compte de ce qu'il existe des moyens de réduire le caractère bruité de la microsimulation.

### **Le caractère stochastique des résultats : comment le contrôler ?**

Recourir à des outils dont le résultat présente une part d'aléa est souvent jugé troublant par les décideurs. Cette préoccupation n'est pas sans paradoxe : elle oublie un peu vite que le même genre d'erreur d'échantillonnage affecte déjà la plupart de nos mesures de la réalité présente : pourquoi exiger d'un modèle de projection qu'il soit plus précis

qu'une mesure de l'existant ? Par ailleurs, la précision apparente des modèles déterministes est en partie illusoire : quelle que soit la méthode de projection, c'est la conception générale du modèle et la qualité de ses hypothèses qui sont le premier déterminant de sa valeur prédictive, et il n'existe évidemment pas de modèles qui échappent à ce problème, qu'ils soient macro ou micro. Il n'en reste pas moins perturbant que, dans le cas de la microsimulation, le risque d'erreur inhérent à tout exercice de prévision se double d'un bruit d'échantillonnage. Même si ce bruit est le prix à payer pour une représentation fine de phénomènes micro qui ne se laissent pas capturer dans un cadre à agents représentatifs, on peut vouloir le minimiser.

Deux premières solutions à peu près équivalentes à ce problème sont l'accroissement de la taille des échantillons simulés, ou la réalisation de simulations multiples dont on présentera les résultats moyens. Mais elles sont coûteuses en temps de calcul, on peut donc chercher à contrôler la variabilité des résultats par d'autres voies. C'est l'objectif des techniques dites de réduction de variance et d'alignement. Elles doivent être différenciées car la seconde soulève davantage de questions que la première. On va en présenter les principes en nous appuyant sur Morrison (2006) ainsi que Li et O'Donoghue (2011)<sup>3</sup>.

Le principe de la réduction de variance est le suivant. Soit la simulation d'un événement particulier à l'intérieur du modèle. Soit  $i = 1, \dots, N$  l'indice désignant les individus et  $p_i$  la probabilité de connaître l'événement pour l'individu  $i$ . À partir de ces probabilités, la méthode de microsimulation par défaut consiste à tirer pour chaque individu un aléa  $u_i$  compris entre zéro et un et à ne faire se réaliser l'événement que pour les individus pour lesquels  $u_i < p_i$ . Le nombre d'événements qu'on obtient dans ce cas sera bien égal à  $\sum p_i$  en moyenne mais avec une variance  $\sum p_i(1-p_i)$ , et donc un coefficient de variation  $(\sum p_i(1-p_i))^{1/2} / \sum p_i \approx 1/(\sum p_i)^{1/2}$  d'autant plus grand que l'événement simulé est rare et la population simulée petite, alors que la simulation déterministe simulerait un nombre de transitions exactement égal à  $\sum p_i$ . L'objectif est de réduire cette variabilité qui n'est pas informative. Il suffit pour cela de contraindre le tirage aléatoire de sorte à ce qu'il donne exactement  $\sum p_i$  résultats positifs. Dit autrement, ce qu'on va demander au tirage aléatoire est de désigner les individus qui vont effectivement connaître l'événement, et non pas leur nombre total qu'on bloque à la valeur  $\sum p_i$ . Pour ce faire, il suffit de mettre en œuvre les techniques classiquement utilisées en sondages pour générer des échantillons de taille fixe à probabilités d'inclusion inégales.

Le calage ou l'alignement sont des généralisations de cette démarche au cas où on veut que le nombre d'événements tirés respecte une autre cible que  $\sum p_i$ . Dans quel cas peut-on souhaiter un tel alignement ? Un exemple est le cas du calage de la microsimulation sur les flux donnés par la projection démographique déterministe classique. Dans cette projection déterministe par la méthode dite « des composants », le flux annuel de naissances ne dépend que des effectifs et des taux de fécondité féminins par âge. De son côté, un modèle de microsimulation fera en général intervenir des déterminants bien plus nombreux, parce qu'il cherche à rendre compte de caractéristiques plus fines de la distribution des tailles de familles selon les caractéristiques des parents. On pourrait évidemment considérer que ce déterminisme plus riche fait la supériorité du modèle de microsimulation et qu'il n'y a pas lieu d'en caler les résultats sur ceux de la méthode traditionnelle. Le calage peut même s'avérer dangereux s'il est seulement utilisé comme moyen de « rattraper » les résultats d'un modèle mal spécifié (Klevmarken, 1998).

3. Voir également une présentation plus détaillée dans Blanchet (2014).

Une précaution minimale est de n'appliquer ce calage que sur un modèle dont on aura d'abord testé le comportement non calé, pour s'assurer que le calage ne jouera qu'un rôle résiduel (Bacon et Pennec, 2009). Mais il existe de nombreuses raisons qui plaident pour un usage raisonné de ce calage. Par exemple, les projections construites par les méthodes traditionnelles font souvent office de projections de référence, et on peut vouloir un modèle de microsimulation qui soit cohérent avec elles, considérant que l'apport de la microsimulation est davantage la prévision des structures fines des ménages que la prospective démographique d'ensemble. Dans ce cas, le modèle de microsimulation ne prétend pas être le mieux placé pour prévoir combien de naissances vont survenir chaque année – il laisse ce soin à la projection de calage – mais il propose une répartition individuelle vraisemblable de ces naissances entre les différents ménages susceptibles d'en avoir une, en respectant les expositions relatives de ces ménages aux probabilités de connaître cet événement.

Plusieurs stratégies sont possibles si on veut mettre en œuvre ce type d'alignement : soit se contenter de minorer ou majorer uniformément les probabilités individuelles avant tirage, au risque que certaines se retrouvent supérieures à un, soit leur appliquer une déformation monotone respectant la contrainte d'avoir tous les  $p_i$  inférieurs à un<sup>4</sup>, soit encore mettre en œuvre une procédure d'alignement dite « par tri » consistant à tirer un aléa  $u_i$  compris entre zéro et un pour chaque individu, puis à les ordonner selon les valeurs croissantes de  $f(u_i)-f(p_i)$  où  $f$  est la transformation *logit*, pour n'en retenir que les  $K$  premiers, ce qui réalise simultanément l'alignement et la réduction de variance.

Plusieurs articles illustrent ces problématiques de calage. Geay et ses coauteurs doivent par exemple redresser leurs probabilités de décès différentielles selon l'état de santé estimées dans les enquêtes *Santé et protection sociale* de l'Irdes de façon à les rendre compatibles avec la mortalité générale des projections de l'Insee, mais sans modifier la structure de leurs probabilités de transition entre la bonne et la mauvaise santé. Des calages peuvent aussi être justifiés dans un modèle de cohorte tel que celui de Courtioux et Lignon qui calent les trajectoires d'activité des générations qu'ils étudient sur les projections de population active de l'Insee.

Des calages sur les projections démographiques et de population active interviennent aussi à des degrés divers dans les modèles de retraite. Tous ces modèles sont par ailleurs amenés à se caler sur les hypothèses de taux de chômage retenues dans les projections du Conseil d'orientation des retraites. Dans ce cas d'espèce, les raisons de l'alignement peuvent paraître plus discutables que l'alignement sur les projections démographiques car aucune méthode n'est meilleure qu'une autre pour prédire les taux de chômage à très long terme : on pourrait donc très bien conserver la dynamique spontanée du chômage microsimulée à partir des probabilités individuelles de transition sur le marché du travail. Dans un tel cas, l'argument qui justifie le calage est plutôt normatif. La vraie question de ces modèles n'est pas de savoir ce que sera le chômage d'ici trente à quarante ans – personne ne peut y répondre, la question est de savoir ce que sera la situation du système de retraite selon que le chômage évoluera de telle ou telle façon. Dès lors que ce niveau futur du chômage est de l'ordre de l'hypothèse plutôt que du résultat, autant fixer cette hypothèse de manière exogène et partagée entre les différents modèles, à charge pour chacun d'entre eux de la traduire au niveau micro-économique, pour bien en déduire les effets individuels et agrégés sur les droits à retraite.

---

4. Pour la conservation des odds ratios associés aux différents  $p_i$ , la déformation à appliquer est de la forme  $p_i = kp_i/(1+(k-1)p_i)$ , comme rappelé dans l'encadré de l'article de Berteau-Rapin et al.

C'est un raisonnement de même nature qui explique la démarche suivie dans l'article de Marbot et Roy. L'exercice que ces auteures présentent est issu de projections réalisées pour l'un des groupes de travail qui avaient été chargés, en 2011, de réfléchir à l'évolution de la prise en charge de la dépendance. Dans ce cadre, deux options polaires étaient possibles pour la projection à long terme des effectifs de personnes dépendantes. La première était la méthode démographique déterministe consistant à appliquer aux projections démographiques des hypothèses raisonnées de prévalence de la dépendance par sexe et âge. La seconde était l'appui exclusif sur la microsimulation consistant à laisser jouer les hypothèses de niveau micro sur les transitions entre autonomie et dépendance. La solution retenue a été un panachage des deux. On a laissé à une projection déterministe classique le soin de déterminer le nombre de personnes dépendantes aux différents horizons. Comme pour le chômage, projeter le nombre global de personnes dépendantes est en effet du domaine de l'hypothèse, puisqu'aucune observation disponible ne permet vraiment de départager entre des scénarios où le calendrier d'entrée en dépendance évolue plus ou moins rapidement que l'espérance de vie globale. Le choix qui a été fait a ainsi été de plutôt exprimer ces hypothèses dans un cadre de modélisation simple, celui de la méthode « matricielle ». En revanche, un modèle de microsimulation était indispensable pour désagréger ces projections macro au niveau individuel, et en tirer des évaluations fines de données telles que l'entourage familial des dépendants, la durée des périodes passées en dépendance, ou les simulations de l'Allocation personnalisée d'autonomie après croisement avec les revenus de retraite : c'est ce qui avait été demandé à Destinie et c'est à cela qu'est consacré l'essentiel de l'article de Marbot et Roy.

Tous ces exemples illustrent bien à quel point il serait hors de propos d'opposer microsimulation dynamique et les méthodes de projection matricielles décrites plus haut, c'est plutôt leur complémentarité qu'il faut relever.

### **Une microsimulation, c'est d'abord des données**

Ces avantages et limites de la méthode ayant été rappelés, il convient maintenant de souligner un point fondamental, le fait que la qualité d'un modèle est faite en large part de la qualité des données sur lequel il s'appuie. La question des bases de données sous-jacentes aux différents modèles occupe ainsi, à juste titre, une large place dans les différents articles qui suivent.

Ce problème n'est évidemment pas propre à la microsimulation. Par exemple, dans le cas d'un modèle macroéconométrique, un travail préliminaire est la constitution de la base de séries macroéconomiques qui serviront à l'estimation des équations du modèle et que celui-ci a ensuite vocation à projeter. Ce travail peut toutefois se borner à prélever l'ensemble de variables requises dans les séries de la comptabilité nationale. Dans un modèle de microsimulation dynamique, l'équivalent est la constitution du fichier de données individuelles initiales que le modèle va ensuite faire vieillir et renouveler. Cette étape peut être assez lourde si elle inclut elle-même une phase importante d'imputation et/ou d'appariement de données. Les problèmes rencontrés sont en fait communs aux modèles statiques et dynamiques.

Un premier cas est celui d'un modèle s'appuyant sur un système d'observation unique, pouvant être constitué d'une seule enquête ou source administrative ou d'un ensemble coordonné de telles sources. Par exemple, le modèle de fiscalité statique Ines et le

modèle Myriade s'appuient sur l'*Enquête Revenus fiscaux et sociaux (ERFS)* de l'Insee, les modèles dynamiques de retraite Prisme et Trajectoire s'appuient respectivement sur les fichiers de cotisants et d'allocataires de la Cnav et sur les *Échantillons interrégimes de cotisants et de retraités (EIC et EIR)* de la Drees appariés au panel *DADS* de l'Insee. Dans ce cas, une partie du travail de préparation des données a été en principe effectuée en amont, lors de la mise en forme des sources considérées. Mais des ajustements supplémentaires peuvent être requis. Ainsi, le dispositif *EIR/EIC* ne couvre que certaines générations, or le modèle vise à projeter un échantillon représentatif de l'ensemble de la population retraitée, ce qui suppose de reconstituer des données pour les générations intermédiaires. Dans le cas du modèle Prisme, les données de gestion de la Cnav doivent être complétées pour les parties de carrière passées ne relevant pas du régime général ou par des données sociodémographiques telles que les dates de naissance des enfants ou l'âge de fin d'études, cette opération de complétion s'appuyant sur l'*EIC* et l'enquête *Étude de l'histoire familiale*. Dans le cas du modèle Autonomix, les dépenses au titre de l'APA issues des remontées administratives doivent être complétées pour estimer l'ensemble du besoin d'aide potentiel des personnes dépendantes. En sus d'être soumise à un ticket modérateur, l'APA est en effet une prestation plafonnée et seul le montant plafonné dit encore « plan d'aide » est connu. Reconstituer le besoin d'aide total est nécessaire si on veut pouvoir simuler des variations du plafonnement de la prise en charge.

Le modèle Destinie est également un modèle s'appuyant sur une source unique, l'enquête *Patrimoine*, mais avec une part de données reconstituées beaucoup plus importante. En premier lieu, l'enquête *Patrimoine* ne donne qu'une partie des informations sur les carrières passées qui sont nécessaires au calcul des droits à retraite. Son apport principal est de fournir des historiques complets des situations sur le marché du travail, à pas annuel. En revanche, en matière de salaires, seul le salaire courant est renseigné dans l'enquête. La séquence des salaires passés doit donc être reconstituée. Elle pourrait l'être par appariement avec d'autres sources comme l'*EIC* mais, dans la version actuelle du modèle on procède plutôt par imputation, à l'aide des mêmes équations de salaires individuels que celles qui sont utilisées en projection, qui font dépendre ce salaire de l'âge, de l'expérience, du genre et du niveau d'éducation, ainsi que du type d'emploi.

La phase d'imputations initiales dans Destinie ne s'arrête pas à cela. Il s'y ajoute une réimputation initiale des montants de retraite pour les individus déjà retraités. La raison initiale de cette imputation était d'éviter tout problème d'incohérence flux/stock au démarrage de la projection : partir de données déclarées en stock et enchaîner sur des pensions reconstituées d'après barème pour les nouveaux liquidants conduit à des évolutions transitoires biaisées à la baisse ou à la hausse selon que les pensions déclarées surestiment ou sous-estiment les pensions vraies découlant des barèmes. Reconstituer les pensions initiales évite ce problème en garantissant l'homogénéité entre ce qui est décrit en stock et ce qui sera projeté en flux. Ceci veut dire que le travail de préparation de la base inclut aussi un travail de rassemblement des règles de calcul passées des droits à retraite, pas seulement de celles qui sont supposées s'appliquer en projection. Ce supplément de travail initial a ensuite une contrepartie intéressante : dès lors que les pensions du stock sont reconstituées, elles peuvent l'être de manière contrefactuelle avec d'autres barèmes que ceux qui ont été réellement appliqués, ce qui permet d'évaluer facilement, par différence, l'effet des réformes passées. Cette possibilité est exploitée par l'article de Dubois et Marino, qui montrent comment les réformes de retraite successives ont rapproché le rendement du système par répartition du niveau qu'il est contraint d'avoir à long terme, égal au taux de croissance de l'économie.

Au total, même dans un modèle s'appuyant sur une base de données réelles, la part d'informations qu'il faut reconstituer peut être assez importante. Mais la part de la reconstitution est encore plus grande avec ce qu'on qualifie de populations « fictives ». Le sens de ce terme doit être bien compris. Il va s'agir de populations auxquelles on va donner des caractéristiques aussi proches que possible d'un ensemble de caractéristiques réelles connues pouvant être issues de diverses enquêtes, mais sans s'appuyer directement sur un fichier de ménages ou d'individus réels. Cette technique peut être utile aussi bien pour des modèles statiques que pour des modèles dynamiques. Dans le numéro joint de la *Revue économique*, c'est ce type de population fictive que Ballot, Goudet et Kant utilisent comme base de départ pour une simulation dynamique du marché du travail français à l'aide du modèle Worksim. Dans ce numéro, c'est à un modèle statique qu'est appliquée la même démarche. Le modèle Taxipp présenté par Antoine Bozio et ses co-auteurs traite une population d'individus qui sont « fictifs » ou « virtuels » mais dont les caractéristiques détaillées, décrites par une quarantaine de variables, sont tirées au hasard de manière à recouper au mieux les données réelles de plusieurs sources statistiques complémentaires : l'enquête *Emploi*, l'enquête *Budget de familles*, l'*ERFS* ainsi que les déclarations fiscales de revenu.

Toujours au chapitre de l'initialisation des données figure la question du choix de la taille de la population simulée. Ce choix dépend à la fois de la taille de la ou des bases de données mobilisées et du type d'utilisation envisagé. Ines utilise les 130 000 individus de l'*ERFS*. Le modèle Myriade de la Cnaf en gère environ 300 000 après « clonage » c'est-à-dire réplique des individus de la même base *ERFS* en fonction des pondérations de l'enquête, ce qui permet ensuite de les traiter de manière équilibrée (Bac *et al.*, 2008). La taille encore plus élevée de l'échantillon de Taxipp s'explique par le souhait de pouvoir zoomer sur des tranches très étroites au sein de la population à très hauts revenus et, s'agissant d'une population fictive, aucune contrainte n'existe quant à cette taille autre que de vitesse de traitement mais ce dernier aspect n'est que secondaire pour un modèle statique : l'échantillon simulé compte 800 000 personnes. Zoomer est également nécessaire lorsqu'on travaille sur une ou quelques générations comme le font Courtioux et Lignon, qui grossissent à environ 150 000 unités le nombre d'individus représentant la génération qu'ils étudient, la génération 1970, un effectif bien supérieur à celui observé dans l'enquête *Emploi* qui est leur source de référence.

L'échantillon de Prisme est encore plus grand : il est obtenu par tirage au vingtième dans les bases de gestion de la Cnav, soit 4 millions d'individus. Cette très grande taille s'explique par le fait que le modèle est mobilisé pour des chiffrages détaillés de mesures portant sur des petites sous-populations, y compris dans une perspective de projections financières à court terme pour lesquelles l'attente de précision est forte. Dans le cas de Destinie, l'objectif est plutôt la simulation des effets tendanciels à long terme de réformes d'assez grande ampleur, ce qui permet de se contenter d'une taille d'échantillon beaucoup plus réduite. Là aussi, cette taille n'est pas forcément celle du fichier de l'enquête : on utilise alternativement des fichiers de taille soit plus petite soit plus grande. Les travaux de mise au point du modèle se sont souvent faits sur un petit fichier au dix-millième de la population française soit seulement six à sept mille individus mais l'échantillon de référence du modèle est un échantillon au millième de 60 000 individus et il en va de même pour Pensipp qui, à ce stade, s'appuie sur les projections de carrières et de trajectoires démographiques fournies par Destinie. L'échantillon de Trajectoire est intermédiaire entre ces deux extrêmes de Destinie et de Prisme, avec 350 000 individus

dans la base *EIC* de départ portés à 1,7 million après imputation des générations non couvertes par cet *EIC*.

Gonfler par clonage l'échantillon de départ est particulièrement utile en microsimulation dynamique. Le clonage des individus de la base initiale ne réduit pas la variabilité des résultats initiaux mais deux clones suivent ensuite des trajectoires différentes ce qui réduit donc la variabilité des résultats en projection. Et le clonage est *a minima* nécessaire pour éviter de gérer des individus ou ménages à pondérations inégales, comme le fait déjà Myriade dans le cadre statique. En microsimulation dynamique, gérer des individus à pondérations inégales poserait notamment problème dans un modèle qui, comme Destinie, simule les ménages en gérant des liens entre individus de base. On voit mal comment gérer de tels liens lorsque les individus ne pèsent pas tous du même poids : par exemple, on ne sait pas quel poids il faudrait donner aux couples formés en réunissant deux personnes de pondérations différentes et par conséquent aux enfants issus de ces couples.

En résumé, la construction de la base de données initiales est une étape essentielle de la construction d'un modèle de microsimulation, avec une grande variété de cas de figure, allant du cas où un fichier de données unique est déjà disponible avec la quasi-totalité des données individuelles que le modèle souhaite projeter, au cas extrême inverse où les données individuelles doivent être totalement reconstituées, en cohérence plus ou moins poussée avec des données agrégées ou semi-agrégées. Entre ces deux cas extrêmes se trouve toute la gamme des cas intermédiaires où les imputations ne portent que sur une partie seulement des données.

## **Paramétrage et validation**

Le travail sur les données de base ne s'arrête cependant pas à la constitution de l'échantillon sur lequel va porter la simulation. Les modèles de microsimulation ont aussi la caractéristique d'être extrêmement gourmands en paramètres. Ceci est source de difficultés à la fois en amont et en aval : en amont parce qu'il faut évaluer ces paramètres et en aval parce qu'il faut s'assurer *a posteriori* que ces paramétrages donnent au modèle un comportement satisfaisant en prévision ou en variantes, et donc prévoir une étape de validation. On va commenter conjointement ces deux problématiques de paramétrage et de validation.

Il y a un premier cas où le paramétrage ne semble pas poser de problèmes statistiques particuliers. Il s'agit de tous les paramètres correspondant aux barèmes des réglementations et des législations que le modèle se propose de simuler. L'absence de problématique statistique ne signifie évidemment pas que la tâche soit légère pour autant, surtout si on veut des modèles rendant compte d'aspects très détaillés des législations. Ceci explique que de nombreux modèles soient directement développés au sein des administrations qui sont les plus proches des thèmes abordés et qui disposent d'une connaissance approfondie de ces barèmes.

Lorsque quasiment tous les paramètres sont de ce type, comme c'est le cas dans les modèles statiques de fiscalité et de transferts sociaux, la validation peut en principe se faire en vérifiant que le modèle restitue bien ce qu'on sait déjà des transferts qu'il est supposé évaluer, en vérifiant en premier lieu que le montant agrégé de ces transferts

est correctement simulé. Même dans ce cadre simple, la démarche peut néanmoins se heurter à des problèmes de cohérence. Soit des problèmes de cohérence entre sources – c’est un des aspects de l’article de Bozio *et al.* – soit en raison des comportements de non-recours à certaines prestations. Cette question est abordée par exemple par Fizzala à propos des allocations versées aux personnes dépendantes. Dès que non-recours il y a, il n’est plus possible de se limiter à de la modélisation comptable, il faut la compléter par au moins un élément de comportement, le fait de demander ou pas une prestation à laquelle on a théoriquement droit. Un article récent d’*Économie et Statistique* a aussi abordé cette prise en compte dans le cas du RSA avec le modèle Myriade de la Cnaf (Domingo et Pucci, 2014).

Les modèles dynamiques comprennent eux-aussi un important volet de paramétrage des législations, présentes ou futures, qui se gère de la même façon. Mais il se double de la question du paramétrage des transitions entre états, avec une gradation dans la difficulté selon le type de comportement que l’on simule.

Un cas qui reste élémentaire est celui où ces probabilités correspondent à des statistiques usuelles, telles que les données des tables de mortalité par sexe et âge, et où le paramétrage consiste à faire des hypothèses sur l’évolution à venir de ces tables. Il n’y a toujours pas de problématique d’estimation ou de validation à proprement parler à ce niveau, pas davantage que dans une projection démographique classique. Des hypothèses d’évolution de la mortalité future ne sont pas « validables » au sens de la validation d’un modèle statistique : tout au plus peut-on porter des jugements *a priori* sur leur vraisemblance, mais ce n’est qu’*a posteriori* qu’on saura si elles auront rendu compte des évolutions démographiques à venir.

Des problématiques d’estimation et de validation ne commencent à apparaître de façon plus marquée que lorsque l’on complexifie le modèle pour y faire jouer des déterminismes démographiques plus sophistiqués que la seule dépendance par rapport au sexe ou à l’âge. Un des intérêts de la microsimulation est, comme on la vu, d’autoriser ces dépendances plus complexes, par exemple des mortalités différentielles par catégorie sociale, niveau de qualification ou état de santé – si le modèle simule ces variables – ou des probabilités de naissances dépendant non seulement de l’âge de la mère, mais aussi du nombre d’enfants déjà nés, du statut d’activité, ou à nouveau de la catégorie sociale ou de la qualification. De même, les transitions sur le marché du travail seront déterminées par des probabilités de passage d’un statut à un autre conditionné également par l’âge, l’ancienneté dans le statut courant, et toujours la catégorie sociale ou la qualification. Rassembler ces éléments représente un travail important pour la qualité du modèle, qui peut s’appuyer à la fois sur la base de données principale du modèle – quand elle a une composante rétrospective – et sur des sources secondaires. Cette étape est importante même lorsque le modèle fait ensuite l’objet de calages car, comme on l’a déjà indiqué, il serait très imprudent de compter sur ces calages pour rectifier des valeurs beaucoup trop mal calibrées de ces paramètres : le remède pourrait être pire que le mal en reportant les défauts d’ajustement sur des aspects de la simulation ne faisant pas l’objet de tels calages.

L’article de Courtioux et Lignon décrit en détail les estimations auxquelles ils ont ainsi dû se livrer pour bien rendre compte de l’homogamie éducative dont ils veulent évaluer l’impact sur la distribution des niveaux de vie. Les articles de Berteau-Rapin *et al.* et Duc *et al.* décrivent pour leur part les démarches suivies pour la simulation des trajectoires

sur le marché du travail. Ces deux modèles s'appuient sur des descriptions fines des transitions entre emploi, inactivité et chômage et, au sein de l'emploi, entre régimes de retraite. Ces transitions sont estimées à partir des données de carrière rétrospectives des fichiers de gestion de la Cnav ou de l'*Échantillon interrégime de cotisants*. Il faut s'assurer que les deux modèles utilisent ces informations de manière satisfaisante, ce qu'on peut faire en leur faisant simuler rétrospectivement les périodes passées. Reconstituer des taux d'activité ou d'emplois corrects et des flux annuels de transition également corrects est évidemment quasi-garanti dès lors que les probabilités de transition ont été directement évaluées à partir de ces flux passés. Mais ceci ne suffit pas : les durées de carrière peuvent être trop peu ou au contraire trop différenciées au niveau micro selon que l'on sous- ou surestime l'hétérogénéité individuelle inobservée ou les phénomènes de dépendance d'état. Appliquer le modèle de façon rétrospective permet de s'assurer que tel n'est pas le cas. Ceci est évidemment très important pour la simulation des droits à retraite : il faut que le modèle produise la bonne proportion de carrières lisses ou heurtées. Les deux articles abordent ce point.

Tous deux sont par ailleurs confrontés, bien évidemment, à la question de la modélisation des comportements de départ en retraite. Prisme et Trajectoire partagent le fait de modéliser le départ en retraite en fonction de la distance à l'âge permettant un départ au taux plein. Cette solution a l'avantage d'être simple et assez cohérente avec les données puisque le taux plein joue encore un rôle de référence dans le système de retraite français. Mais l'atténuation de la décote avant le taux plein, la mise en place et le renforcement de la surcote après ce taux plein ainsi que la suppression de la mise à la retraite d'office à cet âge du taux plein pourraient progressivement lui faire perdre son statut de norme de référence, ce qui constitue un défi pour la modélisation des comportements de départ en retraite. La littérature propose certes des modèles de comportement, plus ou moins élaborés, dans lesquels la date de départ dépend à la fois des barèmes et de paramètres structurels caractérisant les préférences des agents, tels que leur préférence « pour le loisir » ou leur taux d'actualisation subjectif. Mais la mise en œuvre de ces modèles dans un cadre de microsimulation pose une question économétrique particulièrement difficile. En effet, si on veut pleinement tirer parti de l'approche par microsimulation, ce sont non seulement les valeurs moyennes de ces paramètres structurels qu'il faut estimer ou étalonner, ce sont aussi leurs dispersions et leurs corrélations éventuelles, avant de les imputer individuellement aux individus simulés par le modèle. Sans prise en compte de cette diversité, on risque de simuler des réactions trop discontinues aux réformes, avec des individus réagissant en bloc à des changements d'incitations car supposés tous identiques dans leur préférences. Bachelet *et al.* (2011) étaient ainsi amenés à calibrer les variances des paramètres de modèles structurels du départ en retraite pour permettre à Destinie d'alimenter les projections de population active de l'Insee.

Cette hétérogénéité des préférences ou des autres déterminants structurels des choix individuels est essentielle pour expliquer que les effets des réformes ne soient pas de type « tout ou rien » et c'est un apport potentiel des modèles de microsimulation que de permettre d'intégrer ce facteur, mais beaucoup de problèmes techniques sont à résoudre pour que cet apport soit vraiment effectif. On recoupe ce que l'évaluation *ex post* qualifie d'hétérogénéité des effets des traitements et qu'elle s'attache à contrôler et estimer au mieux. Il y a là tout un champ de confrontation potentiellement intéressant entre les méthodes de microsimulation et la microéconométrie de l'évaluation. La qualité d'un modèle de simulation peut s'évaluer par sa capacité à rendre compte des résultats des évaluations *ex post* qui sont disponibles. Dans le cas particulier de la retraite,

le fait que les réformes des retraites passées commencent à se traduire par des hausses significatives de taux d'activité et d'emploi devrait ouvrir des perspectives intéressantes pour ce type de confrontation.

La lecture de l'article de Dubois et Marino suggère pour finir une dernière réflexion sur la question générale de la validation des modèles. Lorsqu'un modèle complexe livre des résultats nouveaux, se pose la question de leur fiabilité. Il est dans ce cas très utile de pouvoir se référer à des ordres de grandeur théoriques : c'est ce que font les deux auteurs en redémontrant comment le taux de rendement moyen d'un système de retraite mature et équilibré est contraint de s'aligner sur le taux de croissance général de l'économie, sous des conditions qui sont assez générales. Ce type de résultat théorique fournit un point d'ancrage contribuant à la validation des résultats centraux du modèle, l'apport de celui-ci étant ensuite de montrer de combien et comment le taux de rendement effectif fluctue autour de cette valeur de référence, en inter- comme en intragénérationnel<sup>5</sup>.

Dit autrement, en présence de gros modèles complexes, il est utile de pouvoir se référer à des modèles plus simples – des « modèles du modèle » – auxquels confier le rôle d'outils de contrôle ou de garde-fous. Le « modèle du modèle » peut être le genre de maquette matricielle ou cellulaire qu'on a discuté plus haut. Mais il peut tout aussi bien s'agir d'un modèle théorique. Là encore, c'est la combinaison d'approches qui permet de tirer le meilleur parti de chacun des outils.

### **Vers davantage d'enrichissement des modèles : modèles bouclés et modèles d'agents**

On conclura cette introduction en évoquant des aspects du programme d'Orcutt dont le traitement reste au mieux implicite dans ce numéro mais qui seront à l'inverse bien représentés dans le numéro associé de la *Revue économique*. Même lorsqu'ils sont dynamiques, on aura pu voir que les modèles présentés dans les articles qui suivent ne répondent qu'à une partie des objectifs initiaux d'Orcutt, qui visait des modèles à comportements totalement endogènes et bouclés. Pour l'essentiel, les modèles présentés sont des modèles plutôt comptables, dans lesquels la modélisation des comportements se réduit le plus souvent à l'application de probabilités exogènes de changement de statut ou d'état, complétées le cas échéant par des endogénéisations de comportement très ciblées – on a discuté l'exemple des comportements de départ en retraite. Idéalement, c'est l'ensemble des comportements des agents qu'il conviendrait d'endogénéiser en réponse aux variations des politiques économiques ou sociales qui sont simulées, et il faut aussi simuler la façon dont ces réactions bouclent entre elles. En particulier, lorsque les questions posées touchent à l'emploi, les modèles de microsimulation qui sont bâtis à partir d'un échantillon de ménages se limitent à une approche d'équilibre partiel qui ne s'intéresse qu'aux réactions de l'offre de travail. Or on a rappelé comment Orcutt avait envisagé d'entrée de jeu la simulation des interactions entre les ménages et les entreprises dans lesquelles travaillent les membres de ces ménages.

---

5. Une autre illustration peut-être trouvée dans Blanchet, Le Minez et Marino (à paraître) qui mobilisent Destinie pour l'évaluation d'un autre indicateur non standard de l'état du système de retraite, le montant de ses engagements implicites, indicateur pour lequel avoir un ordre de grandeur a priori est encore plus difficile que pour le taux de rendement interne.

Pour être tout à fait exact, cette question du bouclage n'est pas totalement absente de ce numéro. On peut même dire qu'elle est d'une certaine façon très présente dans le premier article de Bozio *et al.* car, comme pour tous les travaux analysant les effets des politiques fiscales, ils sont amenés à se poser la question de l'incidence des différents prélèvements : savoir qui supporte en dernier ressort le prélèvement dépend de la façon dont les payeurs apparents peuvent ajuster leur comportements ou leurs prix de sorte à transférer une part plus ou moins grande de la charge de ce prélèvement vers d'autres agents. Sans rentrer eux-mêmes dans une modélisation détaillée des comportements et de l'équilibre des marchés, les auteurs s'appuient sur des résultats classiques d'analyse de la fiscalité en équilibre général pour justifier différentes variantes d'incidence retenues par Taxipp. C'est une façon indirecte d'introduire de l'équilibre général dans un modèle d'apparence purement comptable. Un autre exemple récent de cette démarche pourra être trouvé dans un travail publié récemment par la revue, basé sur le modèle Ines de l'Insee et de la Drees, qui s'était livré à un bilan des effets redistributifs des deux dernières décennies d'évolution du système socio-fiscal (Eidelman, Langumier et Vicard, 2013).

Mais ceci ne répond pas à tous les besoins d'évaluation et tout particulièrement dans les domaines touchant à l'emploi pour lesquels on ne peut clairement pas se contenter d'approches centrées sur l'offre de travail. Que peuvent faire des modèles de microsimulation pour des approches intégrant aussi le côté « demande » de ce marché du travail ? Le numéro associé de la *Revue économique* présentera trois façons de procéder. Une première façon de faire sans pour autant recourir au bouclage est de retourner totalement le projecteur vers ce côté de la demande, ce qui est tout à fait admissible lorsqu'on veut évaluer une politique dont les effets attendus se situent essentiellement de ce côté-là du marché : les micro-unités à simuler ne sont plus les ménages mais les entreprises ou les emplois au sein de ces entreprises. C'est ce qu'Hagneré et Legendre (2016) proposeront pour l'évaluation des effets du CICE, à l'aide du modèle Ambre développé à l'Agence centrale des organismes de sécurité sociale sur la base des *DADS* et des bordereaux récapitulatifs de cotisations des Urssaf. Une telle approche microsimulée du côté demande est aussi celle suivie par la maquette Matis de la DG Trésor (Bock, Lissot et Ozil, 2015).

Si ceci ne suffit pas et si la prise en compte des bouclages est indispensable, on peut l'envisager en articulant un modèle de microsimulation qui reste centré sur les ménages à un modèle d'équilibre général moins désagrégé en charge de simuler l'équilibre de l'ensemble de l'économie. Cette approche a une longue tradition, voir par exemple Cogneau et Robilliard (2007), et le *survey* récent de Cockburn, Savard et Tiberti (2014) : elle sera représentée dans le numéro de la *Revue économique* par une application à l'évaluation *ex ante* de réformes structurelles dans un département d'outre-mer, la Guadeloupe (Mathouraparsad, 2016).

On peut enfin décider de microsimuler l'un comme l'autre des deux côtés du marché du travail. On rejoint alors un autre courant de modélisation, celui des modèles dits d'agents ou ACE pour *agent-based computational economics* qu'on peut voir comme une autre branche de l'héritage d'Orcutt, également nourrie de travaux de type plus sociologiques, consacrés à la simulation des effets d'interaction sociale entre agents à rationalité limitée (Gilbert, 2008). On trouvera des présentations de ce courant dans LeBaron et Tesfatsion (2008), Ballot, Mandel et Vignes (2015) et l'ensemble d'études rassemblées récemment par Gaffard et Napoletano (2012). Des applications françaises de cette approche à la modélisation du marché du travail peuvent être trouvées dans Ballot (2002) puis Barlet

*et al.* (2010) : le premier de ces deux travaux est à la base du nouveau modèle Worksim que Ballot, Goudet et Kant (2016) utiliseront dans le numéro de la *Revue économique* pour évaluer *ex ante* le contrat de génération.

On observe ainsi un début de convergence entre ces modèles ACE et les modèles de microsimulation de facture plus traditionnelle. Ces modèles de microsimulation traditionnels mettent avant tout l'accent sur la représentativité et le réalisme des populations qu'ils simulent, en s'appuyant autant que possible sur des données réelles et une description détaillée des barèmes sociaux et fiscaux qui leurs sont appliqués, sans mettre en avant les questions de comportement ou de bouclage. À l'inverse, les modèles d'agents mettent l'accent sur ces deux derniers points, mais sont longtemps restés appliqués à des populations artificielles fictives, de petite taille, dans des environnements très stylisés. Il s'agit là de différences qui ont vocation à se réduire, avec davantage de prise en compte des comportements et des effets de bouclage dans les modèles de microsimulation et un souci croissant de réalisme des populations simulées dans le cadre des modèles d'agents. Le travail à paraître de Ballot et ses co-auteurs illustre bien cette tendance. Tout ceci peut conduire, progressivement, à la construction de modèles de plus en plus conformes avec le projet initial d'Orcutt. □

---

## BIBLIOGRAPHIE

**Albouy V., Bouton F., Le Minez S. et Pucci M. (2003)**, « Le modèle de microsimulation Ines : un outil d'analyse des politiques sociales et fiscales », *Dossiers Solidarité et Santé* n° 3, Drees.

**Allègre G. et Timbeau X. (2016)**, « Financement du supérieur : les prêts à remboursement contingent sont-ils réellement plus équitables que l'impôt ? Une perspective de cycle de vie », *Revue économique*, à paraître.

**Aubert P., Duc C. et Ducoudré B. (2010)**, « Le modèle Promess : projection "mésos" des âges de cessation d'emploi et de départ à la retraite », *Document de travail*, série *Études et Recherches*, n° 102, Drees.

**Bac C., Legendre F., Lorgnet J.-P. et Marc C. (2008)**, « Recalage de l'enquête Revenus fiscaux et clo-nage des ménages : les deux prérequis à l'utilisation du modèle de microsimulation Myriade », *Courrier des Statistiques*, n° 123, pp. 45-52.

**Bachelet M., Beffy M. et Blanchet D. (2011)**, « Projeter l'impact des réformes des retraites sur l'activité des 55 ans et plus : une comparaison de trois modèles », *Économie et Statistique*, n° 441-442, pp. 123-143.

**Bachelet M., Leduc A. et Marino A. (2014)**, « Les biographies du modèle Destinie 2 : rebasage et projection », *Document de travail Insee/Dese*, n° G2014/01.

**Bacon B. et Pennec S. (2009)**, « Microsimulation, macrosimulation: model validation, linkage and alignment », *International Microsimulation Conference*, Ottawa.

**Ballot G. (2002)**, « Modelling the labour market as an evolving institution: model Artemis », *Journal of Economic Behavior and Organization*, vol. 49, n° 1, pp. 51-77.

**Ballot G., Mandel A. et Vignes A. (2015)**, « Agent-based modeling and economic theory: where are we now? », *Journal of Economic Interaction and Coordination*, vol. 10, n° 2, pp. 199-220.

**Ballot G., Kant J. D. et Goudet O. (2016)**, « Un modèle multi-agents du marché du travail français, outil d'évaluation des politiques de l'emploi. L'exemple du contrat de génération », *Revue économique*, à paraître.

- Bardaji J., Sédillot B. et Walraet E. (2003)**, « Un outil de prospective des retraites : le modèle de micro-simulation Destinie », *Économie et Prévision*, n° 160-161, pp. 193-213.
- Barlet M., Blanchet D. et Le Barbanchon T. (2009)**, « Microsimulation et modèles d'agents : une approche alternative pour l'évaluation des politiques d'emploi », *Économie et Statistique*, n° 429-430, pp. 51-77.
- Bennet R. et Bergmann B. (1974)**, *A microsimulated model of the U.S. economy*, The Johns Hopkins University Press.
- Blanchet D., Buffeteau S., Crenner E. et Le Minez S. (2011)**, « Le modèle de microsimulation Destinie 2 : principales caractéristiques et premiers résultats », *Économie et Statistique*, n° 441-442, pp. 101-121.
- Blanchet D. (2014)**, « La microsimulation dynamique : principes généraux et exemples en langage R », *Document de travail Insee/DMCSI*, n° M2014-1.
- Blanchet D., Le Minez S. et Marino A. (2015)**, « Building and interpreting macro-micro estimates of accrued-to-date pension liabilities: French reforms as a case study », *The Review of Income and Wealth*, à paraître.
- Blanchet D., Bozio A. et Rabaté S. (2016)**, « Quelles options pour réduire la dépendance à la croissance du système de retraite français ? », *Revue économique*, à paraître.
- Bock S., Lissot P. et Ozil S. (2015)**, « Matis : une maquette d'évaluation des effets sur l'emploi de variations de coût du travail », *Document de travail de la DG Trésor*, n° 2015-2.
- Boisnault D. et Fichen A. (2015)**, « Enfants, politique familiale et fiscalité : les transferts du système socio-fiscal aux familles en 2014 », *Trésor Éco*, n° 142.
- Bonnet C. et Mahieu R. (2000)**, « Public pensions in a dynamic microanalytic framework : the case of France », dans Mitton, L., Sutherland, H. and Weeks, M. (éds), *Microsimulation modelling for policy analysis: challenges and innovations*, Cambridge, Cambridge University Press.
- Bourguignon F., Chiappori P.A. et Sastre-Descals J. (1988)**, « Sysiff: a Microsimulation Program of the French Tax-Benefit System », dans A. Atkinson et H. Sutherland, *Tax benefit models*, STICERD-London.
- Bourguignon F. et Spadaro A. (2003)**, « Les modèles de microsimulation dans l'analyse des politiques de redistribution : une brève présentation », *Économie et prévision*, n° 160-161, pp. 231-238.
- Brahimi H., Costes N., Dhont-Peltrault E., Lissot P. et Martin S. (2015)**, « Évaluation des réformes de l'assurance chômage : effets sur les comportements des demandeurs d'emploi et sur la situation financière de l'Unédic - La Maquette IRIs (Impact des réformes de l'indemnisation du chômage) », *Documents de travail de la DG-Trésor*, n° 2015-5.
- Breuil-Genier P. (1998)**, « Les enseignements théoriques et pratiques des microsimulations en économie de la santé », *Économie et Statistique*, n° 315, pp. 73-94.
- Cazenave M.-C., Domingues P. et Vicard A. (2016)**, « Redistribution opérée par "l'avantage HLM" et impact de potentielles réformes : l'apport de la microsimulation », *Revue économique*, à paraître.
- Chanchole M. et Lalanne G. (2012)**, « Photographie du système socio-fiscal et de sa progressivité », *Économie et prévision*, n° 200-201, pp. 19-40.
- Cockburn J., Savard L. et Tiberti L. (2014)**, « Macro-Micro Models », dans *Handbook of Microsimulation Modelling*, O'Donoghue C. (éd.), Emerald Group Publishing Limited, pp. 275-304.
- Cogneau D. et Robilliard A.-S. (2007)**, « Growth, distribution and poverty in Madagascar: Learning from a microsimulation model in a general equilibrium framework », dans *Microsimulation as a tool for*

*the evaluation of public policies : methods and applications*, A. Spadaro (éd.), chap. 3, Fundacion BBVA, Bilbao.

**Courtioux P., Grégoir S. et Houeto D. (2011)**, « Enseignement supérieur et durées de subvention individuelle implicite : une analyse par microsimulation dynamique », *Revue économique*, vol. 62, n° 5, pp. 835-866.

**Debrand T. et Sorasith C. (2010)**, « Apports du modèle de microsimulation Aramis : une analyse des effets redistributifs du plafonnement des restes à charge en ambulatoire », *Questions d'économie de la santé*, Irdes, n° 159.

**Dekkers G. (2015)**, « The simulation properties of microsimulation models with static and dynamic ageing – a brief guide into choosing one type of model over the other », *International Journal of Microsimulation*, vol. 8, n° 1, pp. 97-109.

**Domingo P. et Pucci M. (2014)**, « Impact du non-recours sur l'efficacité du RSA "activité" seul », *Économie et Statistique*, n° 467-468, pp. 117-140.

**Dormont B., Grignon M. et Huber H. (2006)**, « Health expenditure growth: reassessing the threat of ageing », *Health Economics*, vol. 15, n° 9, pp. 947-963.

**Duc C. et Lermecchin H. (2013)**, « Caliper – Un outil de simulation pour le CALCul Interrégimes des PENSIONS de Retraite », *Dossier Solidarité Santé*, n° 37, Drees.

**Duc C., Lequien L., Housset F. et Plouhinec C. (2013)**, « Le modèle de microsimulation Trajectoire (TRAJectoire de Carrières TOus REgimes) », *Document de travail*, série *Sources et méthodes*, n° 40, Drees.

**Duée M. et Rebillard C. (2004)**, « La dépendance des personnes âgées : une projection à long terme », *Document de travail Insee/ Dese*, n° G2004/02.

**Dupont G., Hagneré C. et Touzé V. (2004)**, « Les modèles de microsimulation dynamique dans l'analyse des réformes des systèmes de retraites : une tentative de bilan », *Économie et Prévision*, n° 160-161, pp. 167-192.

**Eidelman A., Langumier F. et Vicard A. (2013)**, « Prélèvements et transferts aux ménages : des canaux redistributifs différents en 1990 et 2010 », *Économie et Statistique*, n° 459, pp. 5-26.

**Eliasson G. (1977)**, « Competition and market processes in a simulated model of the Swedish economy », *American Economic Review*, vol. 67, n° 1, pp. 277-281.

**Favrat A., Prady D. et Tavan C. (2011)**, « Du Smic au niveau de vie : composantes et évolutions », *Économie et Statistique*, n° 448-449, pp. 29-47.

**Favrat A., Marc C et Pucci M. (2015)**, « Les dispositifs sociaux et fiscaux en faveur des familles : quelle compensation du coût des enfants ? », *Économie et Statistique*, n° 478-480, pp. 5-34.

**Gaffard J. L. et Napoletano M. (éds.) (2012)**, *Agent-based models and economic policy*, *Revue de l'OFCE/débats et politiques*, n° 124.

**Geay C., Koubi M. et de Lagasnerie G. (2015)**, « Projections des dépenses de soins de ville, construction d'un module pour Destinie », *Document de travail*, *Insee/Dese*, n° G2015/15.

**Geoffard P.Y. et de Lagasnerie G. (2012)**, « Réformer le système de remboursement pour les soins de ville, une analyse par microsimulation », *Économie et Statistique*, n° 455-456, pp. 89-113.

**Gilbert N. (2008)**, *Agent-based models*, Series: quantitative applications in the social sciences, Sage Publications.

**Goupille J. (2016)**, « Combien ont coûté les réformes de l'impôt sur les successions mises en place en France depuis 2000 ? », *Revue économique*, à paraître.

**Hagneré C. et Legendre F. (2016)**, « Une évaluation ex ante des conséquences du Crédit d'impôt compétitivité emploi (CICE) sur la production et l'emploi », *Revue économique*, à paraître

**Klevmarcken N. A. (1998)**, « Statistical Inference in Microsimulation Models: Incorporating external information », *Working Paper Series, Uppsala University Department of Economics*, n° 1998-20.

**Lachaud C., LARGERON C. et ROCHAIX L. (1998)**, « Franchise sur les soins ambulatoires et équité sociale », *Économie et Statistique*, n° 315, pp. 51-72.

**de Lagasnerie G. (2016)**, « Assurance maladie obligatoire et demande de soins, une analyse par microsimulation », *Revue économique*, à paraître.

**Landais C., Piketty T. et Saez E. (2011)**, *Pour une révolution fiscale : un impôt sur le revenu pour le XXI<sup>e</sup> siècle*, Le Seuil, Coll. « La République des idées ».

**Lardellier R., Legal R., Raynaud D. et Vidal G. (2012)**, « Un outil pour l'étude des dépenses de santé et des restes à charge des ménages : le modèle Omar », *Économie et Statistique*, n° 460, pp. 47-77.

**LeBaron B. et Tesfatsion L. (2008)**, « Modelling Macroeconomics as Open-Ended Dynamic Systems of Interacting Agents », *American Economic Review, Papers and Proceedings*, vol. 98, n° 2, pp. 246-250.

**Legendre F., Lorgnet J.-P. et Thibault F. (2001)**, « La redistribution au bénéfice des familles : l'apport de Myriade », *Recherches et Prévisions*, n° 66, Cnaf.

**Legendre F., Lorgnet J.-P. et Thibault F. (2003)**, « Que peut-on retenir de l'expérience française en matière de microsimulation ? », *Économie et prévision*, n° 160-161, pp. I-XIV.

**Lermechin H., Duc C. et Burricand C. (2011)**, « Présentation et applications de l'outil Caliper (CALcul Interrégimes des Pensions de Retraite) », *Document de travail, série Études et Recherche*, n° 111, Drees.

**Li J. et O'Donoghue C. (2011)**, « Evaluating Alignment Methods in Dynamic Microsimulation Models », 3<sup>rd</sup> International Microsimulation Conference, Stockholm.

**Li J. et O'Donoghue C. (2013)**, « A survey of dynamic microsimulation models: uses, model structure and methodology », *International journal of microsimulation*, vol. 6, n° 2, pp. 3-55.

**Mathouraparsad S. (2016)**, « CloDyn, un modèle EGC microsimulé appliqué à la Guadeloupe », *Revue économique*, à paraître.

**Morand E., Toulemon L., Pennec S., Baggio R. et Billari F. (2010)**, « Demographic modelling: the state of the art », *SustainCity Working Paper*, 2.1a, Ined, Paris.

**Morrison R. (2006)**, « Make it so: event alignment in dynamic microsimulation », *DYNACAN paper*.

**O'Donoghue C. (ed) (2014)**, *Handbook of Microsimulation Modelling*, Emerald Group Publishing Limited.

**Orcutt G. H. (1957)**, « A New Type of Socio-Economic System », *The Review of Economics and Statistics*, vol. 39, n° 2, pp. 116-123.

**Post W. et Van Imhoff E. (1997)**, « Méthodes de microsimulation pour des projections de population », *Population*, vol. 52, n° 4, pp. 889-932.

**Poubelle V. et al. (2006)**, « Prisme, le modèle de la Cnav », *Retraite et société*, n° 48, pp. 202-215.

**Sutherland H. et Figari F. (2013)**, « Euromod: the European Union tax-benefit microsimulation model », *International journal of microsimulation*, vol. 6, n° 1, pp. 4-26.

**Thiébaud S., Barnay T. et Ventelou B. (2013)**, « Aging, chronic conditions and the evolution of future drugs expenditures: A five-year micro simulation from 2004 to 2029 », *Applied Economics*, vol. 45, n° 13, pp. 1663-1672.

**Vernière L. (1990)**, « Les retraites pourront-elles être financées après l'an 2000 ? », *Économie et Statistique*, n° 233, pp. 19-27.

---