

LE MODÈLE DE PRÉVISIONS DE LA CNAF : OÙ EN EST-ON?

Michel GRIGNON*

Au mois de novembre 1991, on peut annoncer que le modèle de prévisions des dépenses de prestations familiales est viable. Rappelons qu'il ne s'agissait pas d'un mince pari puisque la méthode choisie (vieillessement d'un échantillon d'individus) n'avait jamais été employée dans une perspective de prévisions. L'étape suivante est désormais d'améliorer cet outil en gagnant en précision sur certains domaines puis en proposant des procédures d'actualisation simple. Le présent article, après une brève description des techniques mises en oeuvre, fournit l'essentiel des résultats des tests d'adéquation au réel qu'on a fait subir à notre modèle : classiquement, on a fait prévoir au modèle une situation connue par ailleurs et on a pu ainsi simplement comparer prévision et réalisation.

Un précédent numéro de la revue *Recherches et Prévisions* (n°17, De la Gorce, Lefebvre et Grignon, septembre 1989) avait présenté les principes de fonctionnement d'un modèle de prévisions des dépenses de la branche Famille de la Sécurité sociale.

Un modèle de loteries conditionnelles

Rappelons brièvement ses principes : le mécanisme consiste à vieillir d'un an une population de bénéficiaires de prestations familiales. Pour ce faire, on utilise comme population de base l'échantillon tiré au 31 décembre de chaque année dans les bases allocataires des caisses d'Allocations familiales (échantillon dit "au 52ème"). Pour le faire vieillir, on lui applique successivement plusieurs loteries, reflétant les différents événements qui modifient la

situation d'un ménage au regard de la législation des prestations familiales et des aides au logement. Une première loterie détermine ainsi pour chaque famille la réalisation ou la non réalisation de l'événement "naissance d'un enfant au cours de l'année", puis une autre celle de l'événement "séparation du couple parental éventuel", etc.

Le fonctionnement d'une loterie est des plus simples : en fonction de ses caractéristiques socio-démographiques, chaque famille est créditée d'une probabilité $P(i)$ que l'événement "i" se réalise. Ces probabilités, affectées aux familles en fonction de leurs caractéristiques, sont issues d'observations statistiques du passé (issues principalement d'enquêtes de l'Ined, de l'Insee ou de la Cnaf) et de prolongements de tendances. Un cas particulier est constitué par les événements "continus", comme les revenus ou les loyers, pour lesquels l'événement qu'on prévoit est un ensemble de paramètres d'une loi de distribution aléatoire.

Enfin, une fois vieillie la population présente,

* Cnaf - Bureau des Prévisions.

on génère des nouveaux entrants par duplication du dernier flux de nouveaux entrants connu. Le résultat est l'obtention d'un échantillon d'allocataires représentatif de la population future : partant par exemple du dernier échantillon disponible (celui au 31 décembre 1990), on obtiendrait aujourd'hui un échantillon au 31 décembre 1991. En réitérant l'opération, on obtiendrait celui au 31 décembre 1992 etc.

Le double avantage des loteries

L'intérêt d'une telle procédure, certainement plus coûteuse que les procédures économétriques plus classiques, est double : elle offre plus de précision et plus de souplesse.

Plus de précision : dans la mesure où les aides au logement sont en cause et compte tenu, d'une part, de la sensibilité de leurs barèmes aux combinaisons de variables comme le loyer et le revenu et, d'autre part, de la structure complexe de la population bénéficiaire de ces aides, une procédure "atomiste" semble plus sûre que l'utilisation de résumés statistiques sur des sous-populations à laquelle conduisent les procédures économétriques.

Plus de souplesse : en tant qu'aide à la décision tout d'abord, le vieillissement permet de tenir compte simplement de jeux de variantes législatives (il est aisé de modifier les barèmes sur une population future) alors que, dans le cadre d'une prévision reliant directement masses versées et évolution de certaines variables de contrôle, la simulation des variantes supposerait de modifier le coeur même du modèle. En outre, cet aspect modulaire des loteries permet aussi d'utiliser le modèle de prévision comme un test de l'incidence de différents éléments sociaux sur les dépenses de la branche Famille : il suffit pour cela de neutraliser tel ou tel conditionnement en changeant le tableau de probabilités qui lui correspond pour mesurer un effet de ce conditionnement sur la politique familiale. Une telle utilisation a été envisagée pour

tester des hypothèses de microéconomie de la famille.

Les résultats actuels

Outre l'exposé des principes sur lesquels reposait ce modèle, le numéro précité de la revue *Recherches et Prévisions* fournissait de tous premiers résultats permettant de montrer qu'une telle procédure pouvait tourner. Aujourd'hui, l'ensemble des programmes nécessaires à un fonctionnement total du vieillissement sont écrits et il est donc possible de présenter de nouveaux résultats ; il s'agit pour l'instant d'un test du modèle, et on s'est donc attaché à prévoir une population déjà connue afin de mieux mesurer les erreurs afférentes à notre méthode : on a ainsi vieilli d'un an l'échantillon au 31 décembre 1986, puis on a comparé cet échantillon vieilli à celui extrait au 31 décembre 1987.

Les critères retenus pour mesurer l'adéquation de la population vieillie à la population réelle de 1987 reprennent ceux de l'article précédent : distribution par tailles de famille et âge du benjamin (tableau 1), pourcentages de "non en couple" pour chaque croisement taille de famille par âge du benjamin (tableau 2), et pour ce même croisement : pourcentages de femmes chômeuses (tableau 3), revenu annuel moyen de l'homme (tableau 4), aide moyenne au logement perçue (en comptant à 0 l'aide des non bénéficiaires, tableau 5).

En premier lieu, les tableaux montrent que, globalement, la méthode de loteries successives fournit de bons prédicteurs de la réalité. Il subsiste cependant des écarts entre ce qui est simulé et ce qui est observé, écarts qui indiquent les voies à suivre pour améliorer le modèle.

Le tableau 1 se lit simplement : l'avant dernière colonne contient les proportions respectives des différents cas de figure familiaux décrits dans les deux premières (le total de la colonne vaut donc 100 %). La dernière colonne contient les mêmes proportions telles qu'obtenues par la simulation. Ce tableau révèle une différence

de structure liée à l'état actuel des contrôles effectués dans le modèle (les matrices de transition ne sont donc pas en cause) : les familles à moins de deux enfants sont sous représentées dans l'échantillon simulé (l'écart total entre les deux dernières colonnes pour les familles de 0 et 1 enfant est de : moins 15,2 points) car une partie "vidage" du modèle consiste à ne garder dans la population vieillie que les bénéficiaires de l'allocation pour jeune enfant (APJE), des allocations familiales (AF) ou des prestations de logement. On élimine donc à tort les familles de 0 ou 1 enfant présentes dans l'échantillon au seul titre de l'allocation aux adultes handicapés (AAH) ou de l'allocation d'éducation spéciale (AES), voire de l'allocation de parent isolé (API). Autant pour cette dernière prestation il est envisageable de prévoir par le vieillissement toutes les conditions d'ouverture qu'elle requiert, autant les deux premières requièrent des éléments pour l'instant non disponibles sur le handicap; une solution serait de cloner simplement les droits aux prestations autour du handicap. Dans le tableau 2 figurent les pourcentages de parents séparés dans chaque type de famille pour l'échantillon 1987 réel et l'échantillon 1986 vieilli d'un an : la première ligne signifie qu'il y a 60,3 % d'individus non en couple parmi les "familles" allocataires sans enfant de 1987, et 75 % dans l'échantillon 1986 vieilli d'un an. Là encore, les écarts entre ces deux dernières colonnes (qui sont surtout importants pour les familles de moins de deux enfants) reflètent l'effet de structure révélé par le tableau 1 : si on enlève les bénéficiaires d'AAH ou d'AES seules, les chances d'être dans la population allocataire si on a moins de deux enfants sont plus fortes parmi les isolés que parmi les couples.

En revanche, les tableaux 3 et 4 (respectivement, on y lit le taux de femmes au chômage dans chaque configuration familiale et le revenu annuel moyen des hommes) semblent montrer une autre voie d'amélioration possible du modèle, du côté des matrices utilisées dans les loteries. Pour les familles de deux enfants et plus, le revenu de l'homme est toujours un peu surestimé, ce qui semble tenir à la méthode de génération des revenus à partir des paramètres issus de la loterie. Cette surestimation des revenus

explique sans doute la sous-estimation, systématique pour les familles de deux enfants et plus, des montants d'aide au logement perçus.

Ce qui reste à faire

Prévoir le passé : la procédure n'est pas encore totalement bouclée, comme le montrent les tableaux de test. Entre autres, deux lacunes sont particulièrement préjudiciables. En premier lieu, la procédure ne sait pas générer les familles qui, tout en percevant déjà des prestations familiales, vont être de nouvelles entrantes dans le système d'aides au logement. On ne connaît aucune de leurs caractéristiques utiles au calcul d'une aide au logement et on ne peut leur faire subir le sort des nouvelles entrantes pures (duplication du dernier flux connu). Deux solutions sont possibles: soit générer pour toutes les familles hors système d'aide au logement des caractéristiques (loyer, parc) permettant le calcul d'une aide au logement et faire vieillir ces caractéristiques, ou bien générer un flux de nouveaux entrants dans le système d'aide au logement (reproduisant les caractéristiques "logement" du dernier flux connu) qu'il faudra ensuite apparier le mieux possible avec l'échantillon vieilli, de façon à attribuer les caractéristiques logement ainsi générées à des familles de revenu et de tailles adéquates. Pour l'instant, seule la première solution, plus simple, a été explorée et ses résultats sont décevants, car il est difficile de trouver une source qui décrive la population des non bénéficiaires d'aide au logement.

En second lieu, la procédure actuelle néglige une sous-population décisive, à savoir celle des bénéficiaires de l'aide au logement social (ALS), dont l'évolution d'une année sur l'autre obéit à des critères propres, et génère des renouvellements de population important.

Utilisation future du modèle : une fois validée la procédure sur le passé, il restera à passer à la prévision proprement dite, ce qui suppose de savoir faire varier les probabilités conditionnelles dans le temps. Pour ce faire,

on propose le schéma suivant : un bloc de probabilités conditionnelles est établi pour une durée de cinq ans (cela requiert des travaux d'exploitations de nombreuses enquêtes externes), et forme en quelque sorte la structure sur laquelle se déroule le modèle. On ajoute alors à cette structure des effets de tendance, qui visent à simuler l'évolution prévisible des phénomènes sous-jacents aux probabilités : il sera ainsi possible de modifier les probabilités d'arrêt d'activité des femmes à la baisse pour tenir compte d'une tendance générale à l'augmentation de l'activité féminine. D'une certaine manière, on fait ainsi l'hypothèse que, sur une durée assez brève, les évolutions tendanciennes touchent tous les individus identiquement, sans modifier la structure de base.

On superposera enfin à ces effets de conjoncture, sur une base structurelle fixe,

des possibilités de "chocs conjoncturels" pour l'utilisateur : celui-ci pourra, s'il le veut, faire tourner le modèle sous un certain nombre d'hypothèses volontaristes (c'est-à-dire non tendanciennes) du type baisse du taux de chômage.

BIBLIOGRAPHIE

. Développement of a model to examine the relationship between fertility and french family policy. Proceedings of the second International Conference of the American Council for Consumer's Interest (Snowbird, UTAH - august 1990), Robert Mayer, éditeur, 1991.

Tableau 1 - Répartition de la population "réelle" (échantillon au 31.12.1987) et de la population simulée (échantillon 1986 vieilli d'un an) selon les critères "Nombre d'enfants" et "Age du benjamin"

Nombre enfants	Age benjamin	Valeur réelle	Valeur modèle
0	*	12.0	7.3
1	0	3.7	1.5
1	1	3.9	3.3
1	2	3.0	3.2
1	3	8.5	5.3
2	0	3.5	2.7
2	1	3.9	4.4
2	2	3.5	4.0
2	3	32.8	38.7
3	0	1.8	2.3
3	1	1.8	2.0
3	2	1.7	2.1
3	3	12.3	14.4
4	0	0.6	0.5
4	1	0.5	0.6
4	2	0.5	0.6
4	3	2.8	3.4
5 et +	0	0.4	0.3
5 et +	1	0.4	0.5
5 et +	2	0.3	0.4
5 et +	3	1.2	1.4

Tableau 2 - Pour chaque croisement de "Nombre d'enfants" et d'"Age du benjamin", pourcentage de parents ne vivant pas en couple (NB: les ALS sont pour l'instant hors modèle)

Nombre enfants	Age du benjamin	Valeur réelle	Valeur modèle
0	*	60.3	75.0
1	0	15.1	24.8
1	1	15.2	20.2
1	2	17.5	19.6
1	3	45.1	49.3
2	0	5.9	12.8
2	1	6.0	6.2
2	2	7.2	7.7
2	3	11.9	11.8
3	0	4.9	8.4
3	1	5.6	6.3
3	2	6.2	6.8
3	3	10.0	10.0
4	0	6.1	5.6
4	1	7.3	7.7
4	2	6.9	5.6
4	3	10.1	10.1
5 et +	0	5.0	7.2
5 et +	1	5.4	6.1
5 et +	2	9.0	9.3
5 et +	3	11.1	10.6

Tableau 3 - Pour chaque croisement de "Nombre d'enfants" et d'"Age du benjamin", pourcentage de femmes au chômage

Nombre enfants	Age du benjamin	Valeur réelle	Valeur modèle
0	*	9.2	4.7
1	0	11.1	19.1
1	1	12.3	16.2
1	2	11.5	13.3
1	3	10.3	10.4
2	0	7.9	11.0
2	1	8.8	7.5
2	2	6.9	10.4
2	3	3.7	5.6
3	0	3.2	8.1
3	1	3.9	7.9
3	2	4.2	7.1
3	3	3.7	5.4
4	0	4.2	9.2
4	1	2.8	4.1
4	2	1.5	5.6
4	3	3.8	4.6
5 et +	0	4.7	4.8
5 et +	1	2.9	5.3
5 et +	2	1.9	4.2
5 et +	3	2.0	4.3

**Tableau 4 - Pour chaque croisement de "Nombre d'enfants" et d'"Age du benjamin",
revenu annuel moyen de l'homme**

Nombre enfants	Age du benjamin	Valeur réelle	Valeur modèle
0	*	28 601	22 499
1	0	50 871	40 345
1	1	52 922	46 045
1	2	53 498	49 604
1	3	35 036	34 792
2	0	75 458	66 494
2	1	77 189	77 560
2	2	78 134	80 853
2	3	91 542	94 302
3	0	80 912	81 587
3	1	83 623	84 112
3	2	84 526	86 527
3	3	89 216	92 016
4	0	70 715	79 169
4	1	72 209	77 618
4	2	79 826	87 324
4	3	76 380	79 676
5 et +	0	61 143	66 290
5 et +	1	62 651	66 739
5 et +	2	57 923	66 047
5 et +	3	60 985	67 885

**Tableau 5 - Pour chaque croisement de "Nombre d'enfants" et d'"Age du benjamin",
aide moyenne au logement perçue**

Nombre enfants	Age du benjamin	Valeur réelle	Valeur modèle
0	*	552	761
1	0	212	192
1	1	250	240
1	2	255	290
1	3	586	711
2	0	320	286
2	1	367	262
2	2	356	298
2	3	206	184
3	0	620	430
3	1	693	527
3	2	671	568
3	3	469	420
4	0	890	781
4	1	955	796
4	2	945	815
4	3	735	678
5 et +	0	1 147	1 126
5 et +	1	1 226	1 044
5 et +	2	1 192	1 088
5 et +	3	1 098	1 011