

Macro-projection de population vs micro-simulation

1 Introduction

Les méthodes de micro-simulation ont d'abord été développées comme une alternative aux méthodes de macro-projection de population. Les macro-projection de population se donnent comme objet la projection, typiquement pour les cinquante ans à venir, de la pyramide des âges de la population d'un pays. Ces méthodes utilisent la pyramide des âges initiale et recourent à des hypothèses sur les évolutions, respectivement, de la mortalité par âge et de la fécondité des femmes en âge d'avoir des enfants. Notons N_{gt}^f les effectifs des femmes d'âge g de l'année t . Le nombre de femmes, en $t+1$, d'âge $g+1$ se calcule à partir de l'équation :

$$N_{g+1,t+1}^f = ts_{gt}^f \times N_{gt}^f$$

où ts_{gt}^f est le taux de survie des femmes d'âge g de l'année t .

Le nombre total de naissances dépend du nombre de femmes en âge d'avoir des enfants et des taux de fécondité à chaque âge :

$$N_{0,t+1} = \sum_{g=15}^{45} tf_{gt}^f \times N_{gt}^f$$

où tf_{gt}^f est le taux de fécondité des femmes d'âge g de l'année t – en supposant que seules les femmes de 15 à 45 sont en âge d'avoir des enfants.

L'exercice proposé ici a deux buts

- réaliser une projection de population pour la FRANCE simple mais qui serait réaliste ;
- mettre en œuvre, comme méthode alternative, une micro-simulation.

2 Une macro-projection pour la France

Pour faire simple, nous allons agréger la pyramide des âges par tranches d'âge vicennal. Nous distinguons ainsi les femmes âgées de 0 à 19 ans, de 20 à 39 ans, etc. De même pour les hommes.

2.1 Les données Les données nécessaires sont disponibles par l'intermédiaire du site Web de l'INÉD – l'Institut national d'études démographiques ; on obtient, sur le champ « France entière » au premier janvier 2009, après consolidation par tranche d'âge vicennal, les chiffres portés dans le tableau ci-après.

Num.	Tranche d'âge vicennal	Femmes	Hommes
1	0 - 19 ans	7 779 318	8 137 171
2	20 - 39 ans	8 356 902	8 300 920
3	40 - 59 ans	8 912 230	8 540 872
4	60 - 79 ans	5 941 230	5 046 948
5	80 - 100 ans	2 181 379	1 092 601
6	+ de 100 ans	12 302	1 609

Source : INSEE via le site Web de l'INÉD.

Il naît plus de garçons que de filles ; les femmes vivent cependant beaucoup plus âgées que les hommes. La faible fécondité des 25 dernières années fait que les effectifs de la tranche vicennale 1 sont plus réduits que ceux de la tranche 2. Cette faible fécondité provient d'abord d'un recul de l'âge à la maternité.

On trouve aussi depuis le site de l'INÉD une table de mortalité construite à partir des données des années 2004-2006. Cette table donne notamment les taux de survie à différents âges, par sexe. On a ainsi, exprimés en pour 100 000, les chiffres suivants.

Age	Femmes	Hommes
10	99 560	99 445
30	99 113	98 270
50	96 959	93 809
70	87 907	74 753
90	37 116	17 481

Ces taux sont inconditionnels ; ils ne donnent pas la probabilité de survie une année sachant que l'on a déjà survécu toutes les années précédentes.

Pour simplifier l'exercice, on va supposer que seules les femmes de la deuxième tranche vicennale sont fécondes ; on retient un taux de natalité de 2 enfants par femme.

2.2 La macro-projection obtenue grâce à Excel Le plus simple est de recourir à un tableur pour réaliser la macro-projection, et ce d'autant plus que les données ont été consolidées par tranche d'âge vicennale. Dans la table 1 figure la feuille de calcul ; le classeur Excel peut aussi être téléchargé depuis <http://fj.legendre.free.fr> pour obtenir le détail des formules de calcul.

Pour la population des plus de 100 ans, j'ai retenu des probabilités de survie arbitraires : 15% pour les femmes et 5% pour les hommes pour bien montrer que ces exercices de projection dépendent des hypothèses retenues sur l'évolution de la mortalité

TABLE 1 – La feuille de calcul de la macro-projection

2009 Tranche d'âge	Population		Survie	
	Femmes	Hommes	Femmes	Hommes
0-19 ans	7779318	8137171	99560	99445
20-39 ans	8356902	8300920	99113	98270
40-59 ans	8912230	8540872	96959	93809
60-79 ans	5941230	5046948	87907	74753
80-99 ans	2181379	1092601	37116	17481
+ de 100 ans	12302	1609	Total	64303482

2029 Tranche d'âge	Population		Survie	
	Femmes	Hommes	Femmes	Hommes
0-19 ans	8133069	8497368	99560	99445
20-39 ans	7744391	8041026	99113	98270
40-59 ans	8175283	7924097	96959	93809
60-79 ans	8080193	6805912	87907	74753
80-99 ans	2508500	1180230	37116	17481
+ de 100 ans	327207	54630	Total	67471904

2049 Tranche d'âge	Population		Survie	
	Femmes	Hommes	Femmes	Hommes
0-19 ans	7541417	7870113	99560	99445
20-39 ans	8096553	8396967	99113	98270
40-59 ans	7576084	7676001	96959	93809
60-79 ans	7412047	6314426	87907	74753
80-99 ans	3411610	1591564	37116	17481
+ de 100 ans	376275	59011	Total	66322067

(et, implicitement, sur les progrès de la médecine et de l'attention que la société porte aux personnes âgées).

3 La micro-simulation obtenue grâce au système Sas

La première étape DATA (lignes 1 à 14) permet d'acquérir les données de la pyramide des âges. Il est plus simple de copier les chiffres par un copier/coller dans le programme Sas que d'utiliser la PROC import quand la feuille de calcul ne s'y prête pas bien. Noter l'utilisation de la variable automatique _N_ pour renseigner la variable tav (la tranche d'âge vicennale).

La deuxième étape DATA (lignes 18 à 27) sert à lire les taux de survie. Ces taux sont des paramètres; il importe ainsi de les placer dans une table Sas qui ne comporte qu'une seule ligne. C'est l'objet de la troisième étape DATA (lignes 28 à 36).

L'enquête est engendrée de manière artificielle en utilisant les effectifs de la pyramide des âges initiale. C'est l'objet de l'étape DATA des lignes 37 à 43. Noter que c'est étonnamment simple. La micro-simulation est programmée de la ligne 44 à la ligne 105.

```

1  DATA effectifs ;
2      INPUT nbf nbh ;
3      tav = _N_ ;
4      sexe = 'F' ; nombre = nbf ; OUTPUT ;
5      sexe = 'H' ; nombre = nbh ; OUTPUT ;
6      CARDS ;
7      7779318 8137171
8      8356902 8300920
9      8912230 8540872
10     5941230 5046948
11     2181379 1092601
12     12302 1609
13 ;
14 RUN ;
15 %LET proportion_filles = 7779318 / (7779318 + 8137171) ;
16 %LET survie_femmes_centenaires = .15 ;
17 %LET survie_hommes_centenaires = .05 ;
18 DATA survie ;
19     INPUT ps_f ps_h ;
20     CARDS ;
21     99560 99445
22     99113 98270
23     96959 93809
24     87907 74753
25     37116 17481
26 ;
27 RUN ;
28 DATA survie ;
29     SET survie END=fin ;
30     RETAIN psf1-psf5 psh1-psh5 ;
31     ARRAY psf psf1-psf5 ; ARRAY psh psh1-psh5 ;
32     psf[_N_] = ps_f ;
33     psh[_N_] = ps_h ;
34     IF fin THEN OUTPUT ;
35     KEEP psf1-psf5 psh1-psh5 ;
36 RUN ;
37 DATA individus (KEEP = sexe tav poids) ;
38     SET effectifs ;
39     poids = 1000 ;
40     DO i = 1 TO nombre/poids ;
41         OUTPUT ;

```

```

42 END ;
43 RUN ;
44 DATA individus_2029 (KEEP = sexe tav poids) ;
45 IF _N_ EQ 1 THEN SET survie ;
46 ARRAY psf psf1-psf5 ; ARRAY psh psh1-psh5 ;
47 SET individus ;
48 IF (sexe EQ 'F') AND (tav EQ 2) THEN DO ;
49   tav = 1 ; /* Tranche d-âge vicennal de l-enfant. */
50   /* 20 % des femmes ont un enfant ; 80 % deux enfants ;
51     20 % trois enfants. 2 = 20%*1 + 60%*2 + 20%*3. */
52   /* Un peu moins d-une chance sur deux d-avoir une fille. */
53   IF RANUNI(12345678) < &proportion_filles THEN DO ;
54     sexe = 'F' ; proba_survie = psf{1} / 100000 ;
55   END ;
56   ELSE DO ;
57     sexe = 'H' ; proba_survie = psh{1} / 100000 ;
58   END ;
59   /* Le premier enfant, s-il survit. */
60   IF RANUNI(12345678) < proba_survie THEN
61     OUTPUT ;
62   IF RANUNI(12345678) < (.6+.2) THEN DO ;
63     IF RANUNI(12345678) < &proportion_filles THEN DO ;
64       sexe = 'F' ; proba_survie = psf{1} / 100000 ;
65     END ;
66     ELSE DO ;
67       sexe = 'H' ; proba_survie = psh{1} / 100000 ;
68     END ;
69     /* Le deuxième enfant, s-il survit. */
70     IF RANUNI(12345678) < proba_survie THEN
71       OUTPUT ;
72     IF RANUNI(12345678) < .2/(.6+.2) THEN DO ;
73       IF RANUNI(12345678) < &proportion_filles THEN DO ;
74         sexe = 'F' ; proba_survie = psf{1} / 100000 ;
75       END ;
76       ELSE DO ;
77         sexe = 'H' ; proba_survie = psh{1} / 100000 ;
78       END ;
79       /* Le troisième enfant, s-il survit. */
80       IF RANUNI(12345678) < proba_survie THEN
81         OUTPUT ;
82       END ;
83     END ;
84     tav = 2 ; /* On restaure les valeurs des variables de la mère. */
85     sexe = 'F' ;
86   END ;

```

```

87 IF tav NE 6 THEN DO ;
88   IF tav EQ 5 THEN DO ;
89     IF sexe EQ 'F' THEN
90       proba_survie = &survie_femmes_centenaires ;
91     ELSE
92       proba_survie = &survie_hommes_centenaires ;
93   END ;
94   ELSE DO ;
95     IF sexe EQ 'F' THEN
96       proba_survie = psf{tav+1} / psf{tav} ;
97     ELSE
98       proba_survie = psh{tav+1} / psh{tav} ;
99   END ;
100   IF RANUNI(12345678) < proba_survie THEN DO ;
101     tav = tav + 1 ;
102   OUTPUT ;
103   END ;
104   END ;
105 RUN ;
106 PROC freq DATA = individus ;
107   WEIGHT poids ;
108   TABLE sexe*tav ;
109 RUN ;
110 PROC freq DATA = individus_2029 ;
111   WEIGHT poids ;
112   TABLE sexe*tav ;
113 RUN ;

```

Les résultats agrégés sont obtenus par une PROC freq; d'abord sur le fichier initial ensuite sur le fichier qui résulte de la micro-simulation. Les différences entre les résultats agrégés de la micro-simulation et les résultats de la macro-projection relèvent de l'aléa inhérent de la micro-simulation.

Table of sexe by tav							
sexe	tav						
	1	2	3	4	5	6	Total
F	7779000	8356000	8912000	5941000	2181000	12000	3.318E7
	12.10	13.00	13.86	9.24	3.39	0.02	51.61
	23.44	25.18	26.86	17.90	6.57	0.04	

13		48.88	50.17	51.07	54.07	66.64	92.31	
14								
15	H	8137000	8300000	8540000	5046000	1092000	1000	3.112E7
16		12.66	12.91	13.28	7.85	1.70	0.00	48.39
17		26.15	26.67	27.45	16.22	3.51	0.00	
18		51.12	49.83	48.93	45.93	33.36	7.69	
19								
20	Total	1.592E7	1.666E7	1.745E7	1.099E7	3273000	13000	6.43E7
21		24.75	25.90	27.14	17.09	5.09	0.02	100.00

1 Table of sexe by tav

2 sexe tav

3

4 Frequency |

5 Percent |

6 Row Pct |

7 Col Pct | 1 | 2 | 3 | 4 | 5 | 6 | Total

8 -----+-----+-----+-----+-----+-----+-----+-----

9 F |8125000|7734000|8179000|8062000|2497000|331000|3.493E7

10 | 12.04| 11.46| 12.12| 11.95| 3.70| 0.49| 51.77

11 | 23.26| 22.14| 23.42| 23.08| 7.15| 0.95|

12 | 49.10| 48.98| 50.81| 53.98| 67.40| 84.65|

13 -----+-----+-----+-----+-----+-----+-----+-----

14 H |8423000|8056000|7919000|6873000|1208000| 60000|3.254E7

15 | 12.48| 11.94| 11.74| 10.19| 1.79| 0.09| 48.23

16 | 25.89| 24.76| 24.34| 21.12| 3.71| 0.18|

17 | 50.90| 51.02| 49.19| 46.02| 32.60| 15.35|

18 -----+-----+-----+-----+-----+-----+-----+-----

19 Total | 1.655E7| 1.579E7| 1.61E7| 1.494E7| 3705000| 391000| 6.747E7

20 | 24.53| 23.40| 23.86| 22.14| 5.49| 0.58| 100.00