



Mémoire présenté
devant l'Institut de Science Financière et d'Assurances
pour l'obtention du diplôme d'Actuaire de l'Université de Lyon
le 28 novembre 2011

Par : Alexandre CALLE

Titre : Projection de population par microsimulation

Confidentialité : NON OUI (Durée : 1 an 2 ans)

Membre du jury de l'Institut des Actuaires

M. Christian FETTIG

Entreprise :

SPAC Actuaires

Membres du jury I.S.F.A.

M. Jean Claude AUGROS

M. Alexis BIENVENÛE

M. Areski COUSIN

Mme Diana DOROBANTU

Mme Anne EYRAUD-LOISEL

M. Nicolas LEBOISNE

M. Stéphane LOISEL

Mlle Esterina MASIELLO

Mme Véronique MAUME-DESCHAMPS

M. Frédéric PLANCHET

M. François QUITTARD-PINON

Mme Béatrice REY-FOURNIER

M. Pierre RIBEREAU

M. Christian-Yann ROBERT

M. Didier RULLIERE

M. Pierre THEROND

Directeur de mémoire en entreprise :

Olivier Seguin

Invité :

**Autorisation de mise en ligne sur
un site de diffusion de documents
actuariels (après expiration de
l'éventuel délai de confidentialité)**

Signature du responsable entreprise

Signature du candidat

Secrétariat

Mme Marie-Claude MOUCHON

Bibliothèque :

Mme Michèle SONNIER

Mémoire d'Actuaire (M2 PRO SAF)

Projection de population par microsimulation

Alexandre Calle



Résumé

Mots clefs : microsimulation, modélisation de comportement, projection de population, Monte Carlo, aléa,

Une entité envisage de modifier les règles actuelles d'acquisition d'un avantage par mariage. Actuellement, seules les femmes « étrangères » mariées à un homme issu de la population cible peuvent demander l'avantage après cinq ans de mariage. L'entité envisage d'étendre cette possibilité aux deux sexes après dix ans de mariage et souhaiterait tester l'impact sur l'évolution des effectifs à long terme.

Les méthodes de microsimulation, travaillant au niveau individuel plutôt que collectif, permettent de prendre en compte un grand éventail de paramètres afin de modéliser les comportements humains, très hétérogènes au sein d'une population. Les résultats d'un modèle de microsimulation étant aléatoires, cela permet de tester leur sensibilité. La microsimulation est donc une méthode particulièrement adaptée à cette étude qui nécessite un nombre important d'événements démographiques incertains. Le principe est de soumettre à chaque individu l'aléa de ces événements créant ainsi une multitude de biographies individuelles.

Divers scénarios ont été testés concernant le comportement des « étrangers » vis-à-vis de la demande d'avantage par mariage. Un premier scénario considère que le comportement des hommes sera identique à celui des femmes observé jusqu'à présent. Un second scénario prend l'hypothèse que toutes les personnes mariées depuis plus de 10 ans demanderont l'avantage dans les trois ans qui suivent.

La structure des données de base pour l'étude a elle aussi été testée. Un scénario avec une population uniforme pour chaque génération, corrigeant notamment le « creux » significatif des individus autour de 30 ans de la population d'origine, a été modélisé. D'autres scénarios portant sur des nombres d'entrées différents ont aussi été produits afin de mesurer la sensibilité de cette variable sur le nombre d'individus aux différents horizons (2020, 2030 et 2040).

Les résultats mettent en relief un effet « d'aubaine » dans tous les scénarios avec la réglementation envisagée. En effet, un pic de croissance est observé pour chaque simulation dans les trois ans puis un rythme de croisière s'installe avec un nombre de demande d'avantage par mariage supérieur à celui d'aujourd'hui sans toutefois la doubler.

Abstract

Keywords: microsimulation, modelling behaviour, projection, Monte Carlo, hazard

An entity is considering amending the current rules for acquiring an advantage by marriage. Currently, only women can acquire it after five years of marriage. The entity plans to extend this opportunity to both sexes after ten years of marriage and would like to test the impact on the population.

Microsimulation methods, based on individual rather than group of people, can accommodate a wide range of parameters to model human behaviour, very heterogeneous within a population. The results of microsimulation models are random, it can test their sensitivity. Microsimulation is a method particularly suitable for this study which requires a large number of demographic uncertain events. The idea is to submit to each individual the hazard of these events, creating a multitude of individual biographies.

Various scenarios have been tested concerning the behaviour of “foreign” with the demand for advantage by marriage. The first scenario considers that men’s behaviour will be identical to that of women observed so far. A second scenario is the assumption that all married people over 10 years applying for the advantage in the following three years.

The structure of the data for the study was also tested. A scenario with a uniform population for each generation, including correcting the "hollow" significant people around 30 years of the original population, was modelled. Other scenarios involving different numbers of entries were also produced to measure the sensitivity of this variable on the number of people with different backgrounds (2020, 2030 and 2040).

The results highlight an effect of "windfall" in all scenarios with the proposed legislation. Indeed, a growth peak is observed for each simulation within three years and then moved a stride with a higher number of advantage asking by marriage than today without be the double.

Remerciements

Je voudrais utiliser ces quelques lignes pour remercier les personnes qui ont contribué de près ou de loin à ce mémoire et plus généralement celles qui m'ont encouragé dans ce travail.

Je tiens à remercier sincèrement toute l'équipe de SPAC Actuaires qui m'a accompagné pendant toute l'écriture de ce mémoire, durant l'alternance et encore maintenant, plus particulièrement Brigitte Ecary et Olivier Seguin. Leurs conseils, relectures et critiques m'ont permis de terminer ce travail dans d'excellentes conditions.

Je suis reconnaissant à l'égard de monsieur Alexis Bienvenüe, mon tuteur de mémoire, ainsi que de l'ensemble du personnel de l'ISFA.

Mes remerciements s'adressent aussi à Didier Blanchet, chef du département des études économiques d'ensemble à l'INSEE qui a su se rendre disponible pour répondre à mes interrogations.

Enfin je tiens à remercier toutes les personnes qui ont eu une influence sur ce mémoire. Merci à Emilie, aux Gimmicks, à ma famille et aux nombreuses personnes qui ont pu m'aiguiller lors de ce travail.

Table des matières

Introduction	9
Contexte	11
Partie I : La microsimulation, principes et utilisations	13
A. La méthode des composantes	16
1. Principe	16
2. La méthode des composantes dans notre étude.....	20
B. Caractéristiques de la microsimulation	22
1. Les différents modèles de microsimulation.....	22
2. Atouts et limites de la microsimulation	25
3. L'aléa en microsimulation	28
4. La hiérarchisation des risques	31
5. Les dernières avancées en microsimulation.....	32
Partie II : Travaux réalisés et 1ers résultats	35
A. La construction de la base de données initiale	35
1. Les retraitements effectués.....	36
2. La description de la base de données finale	38
B. Le calibrage des paramètres de projections	40
1. La natalité	40
2. La mortalité	43
3. La nuptialité.....	44
4. La divortialité	46
5. Les entrées.....	47
6. Les taux de demande d'avantage par mariage.....	48

C. Programmation du modèle et premiers résultats	49
1. Présentation du modèle de microsimulation	49
2. Premiers résultats	51
Partie III : Amélioration du modèle	54
1. Modèle de taux à un facteur	54
2. Le modèle d'Ornstein-Uhlenbeck	56
2. Le générateur aléatoire utilisé	60
Partie III : Résultats des projections à réglementations distinctes	61
A. Les résultats des projections à réglementation actuelle	62
B. Les résultats des projections à réglementation envisagée	67
1. Scénario 1	67
2. Scénario 2	73
Conclusion	77
Annexe 1	78
Annexe 2	80
A. Natalité	80
B. Nuptialité	81
C. Divortialité	82
Annexe 3 : Résultats complémentaires	83
A. Projections après correction du creux de la population autour de 30 ans	83
1. Projections à réglementation actuelle (n/2 entrées)	85
2. Projections à réglementation envisagée (n/2 entrées)	87
a. Scénario 1	87
b. Scénario 2	89
B. Projections sans prise en compte des entrées	92

1. Projections à réglementation inchangée.....	92
2. Projections à réglementation envisagée.....	94
a. Scénario 1	94
b. Scénario 2.....	96
C. Projections en considérant la moyenne historique des entrées par an.....	99
1. Projections à réglementation inchangée.....	99
2. Projections à réglementation envisagée.....	101
a. Scénario 1	101
b. Scénario 2.....	103
Bibliographie	106

Introduction

Les comportements humains ont toujours évolué au fil du temps en fonction de plusieurs facteurs : changements législatifs, évolution sociétale, acquisitions d'expériences... Aujourd'hui, en France et ailleurs, l'analyse de ces évolutions comportementales fait de plus en plus l'objet d'études qui tentent de prévoir, avec les données d'aujourd'hui, les comportements de demain.

Les moyens informatiques ayant énormément évolué durant ces dernières années, de nouvelles méthodes de projections voient le jour. La microsimulation en se basant au niveau des individus ou des ménages et non plus au niveau d'un groupe d'individus, permet de prendre en compte un éventail beaucoup plus large de paramètres lors des projections.

En particulier, les comportements humains, très hétérogènes au sein d'une population, peuvent maintenant être modélisés avec une précision accrue.

Il est donc devenu possible de mesurer les effets probables des réformes pour les entreprises et les politiques avant qu'elles ne soient appliquées dans le but de mieux se prémunir contre les risques inhérents à ces changements et de mieux évaluer les conséquences de ces derniers.

Cette étude étant confidentielle, bon nombre de résultats et précisions concernant le contexte de cette étude sont volontairement omis afin de respecter la confidentialité qui incombe au cabinet. Une annexe confidentielle, sans aucune limitation de durée, est à la disposition du jury de ce mémoire.

Nous allons projeter une population particulière aux horizons 2020, 2030 et 2040. La population étudiée est composée d'hommes et de femmes de tous âges dont chacun des membres s'est vu octroyer un certain avantage au cours de sa vie par la même entité. Cette entité octroie aux femmes « étrangères » un certain avantage après 5 ans de mariage avec un homme issu de la population de base. A ce jour, les hommes « étrangers » mariés à une femme issue de la population de base ne peuvent prétendre à ce dernier. Il est envisagé d'étendre cet avantage à tous après 10 ans de mariage. Avant de modifier ces règles, nous allons procéder à une projection des effectifs probables sur les prochaines années, dans le

cadre de la réglementation actuelle, et dans le cadre des modifications envisagées selon plusieurs scénarios comportementaux.

L'étude réalisée a donc pour but d'appréhender l'évolution de cette population suite à un changement de règle relatif aux modes d'acquisition de l'avantage.

La méthode employée dans ce cadre est la microsimulation : appliquée à notre étude elle a pour principe de projeter la population d'aujourd'hui jusqu'à divers horizons avec différents jeux d'hypothèses. Cette méthode simulera virtuellement chaque vie de façon individuelle et déterminera via des tirages aléatoires de Monte Carlo l'évolution de chaque caractéristique pertinente relative à l'étude.

Dans le cadre de ce travail, nous nous intéresserons tout d'abord à présenter la microsimulation et à justifier son utilisation pour ce type d'étude. La partie suivante présentera le modèle de microsimulation construit, les retraitements des données qui ont été en notre possession ainsi que le calibrage des paramètres de simulation. Les résultats des simulations dites « centrales » ainsi que les tests de sensibilité seront détaillés dans la troisième partie de ce mémoire. La quatrième partie exposera les résultats de différents scénarios.

Contexte

Cette étude intervient dans le cadre de l'activité de SPAC actuaires. En conséquence certaines informations ou résultats de l'étude seront volontairement omis afin de respecter la confidentialité qui incombe au cabinet. Notamment, les graphiques sont données sans échelles.

Cette étude a pour but de projeter une population cible aux horizons 2020, 2030 et 2040 selon deux scénarios, l'un qui reprend la règle actuelle et l'autre qui se place dans l'hypothèse de la réforme envisagée.

Aujourd'hui, seules les femmes « étrangères » mariées à un homme de la population cible peuvent prétendre à l'avantage passé un délai de 5 ans. Les hommes « étrangers » mariés à une femme de la population cible ne peuvent prétendre à ce jour à cet avantage. Par souci d'équité il est donc envisagé d'étendre cette possibilité à tous les sexes après un délai de 10 ans de mariage.

Dès qu'une personne acquiert cet avantage, elle est immédiatement considérée comme faisant partie intégrante de la population cible et est ajoutée aux effectifs.

Avant de réformer selon les termes énoncés ci-dessus, l'entité souhaiterait avoir une projection des effectifs probables : à règle actuelle et dans celle envisagée. En effet, la population étant restreinte, une augmentation significative des effectifs d'ayants droit à l'avantage doit être anticipée afin de faire face tant au niveau infrastructurel qu'au niveau administratif.

Le scénario à règle actuelle ne possède aucune variante. En effet, dans les simulations les comportements actuels des femmes sont connus en ce qui concerne les demandes d'octroi d'avantage et nous avons donc tenté de les reproduire dans nos simulations.

Le scénario à réglementation envisagée possède deux sous variantes : la première part du principe que les hommes auront le même comportement que les femmes et donc nous avons calé les comportements théoriques des hommes sur celui des femmes observé aujourd'hui (scénario 1).

La seconde est dite « catastrophe » : chaque homme « étranger » marié avec une femme issue de la population cible depuis au moins 10 ans demandera l'avantage (avec une fenêtre de demande de 3 ans, scénario 2).

Il est important de préciser qu'une personne qui obtient l'avantage et ne fait pas partie de la population de base dès sa naissance, ne peut en aucun cas transférer cet avantage à une tierce personne.

Le choix de la méthode, la microsimulation, s'est très vite imposé à nous. En effet, la population étant restreinte et le fait que nous souhaitions travailler avec des intervalles de confiance directement disponibles ont fait que nous avons opté pour cette dernière. Cependant, comme nous allons le montrer dans la partie suivante, le choix des méthodes de projections usuelles aurait été très vite problématique dans le cadre de notre étude.

Partie I : La microsimulation, principes et utilisations

« L'entre-deux-guerres a vu l'apparition, la diffusion et l'adoption universelle de la méthode dite « des composantes » pour le calcul des projections de population. On a fait jouer l'interaction entre la structure par sexe et âge de la population et les taux par sexe et âge des composantes du mouvement démographique (la mortalité, la fécondité et la migration) pour indiquer l'évolution future la plus vraisemblable des populations nationales. Mais, dès le début, cette approche fondée sur l'analyse démographique a été concurrencée par une résurgence de la théorie selon laquelle l'effectif futur de la population est déterminé par une loi. Cette croyance s'appuyait sur le (re)découverte d'un modèle homéostatique, la « loi logistique » de la croissance démographique. L'approche logistique des projections de population a été développée et préconisée par le généticien américain Raymond Pearl dans les années 1920. Elle a alors supplanté la loi malthusienne de la croissance géométrique, qui avait régné au XIXe siècle. La décennie des années 1920 vit donc s'affronter l'approche « par la loi logistique » et l'approche « par l'analyse démographique ». [...] La discussion sur les méthodes de projection de population opposa tout d'abord la biologie et la démographie. La controverse se développa dans des congrès, des articles et des ouvrages, aux frontières du domaine qui avait vu naître les innovations techniques. La méthode des composantes trouva aisément sa place dans le champ de la planification. Elle permettait une compréhension précise des facteurs qui expliquent la dynamique de la population, et elle fournissait un profil détaillé de la structure par sexe et âge de la population future. » (Henk, 2002)

Jusqu'à présent la principale méthode de projection de population était la méthode des composantes. Le principe repose sur le fait que la population initiale sert de base de données afin de calculer les différents paramètres servant à projeter cette dernière : taux de fécondité, de mortalité etc. Ces projections sont réitérées pour chaque intervalle de temps considéré où la population issue de la projection N-1 sert de population de base pour la projection N.

Cette méthode nous permet de travailler au niveau global de la population considérée et si pour 10 000 femmes de même caractéristiques il a été estimé un taux de fécondité de 2% annuel alors la projection sur un intervalle d'un an nous donnera exactement 200 naissances.

La microsimulation étudie l'évolution d'un système en utilisant les caractéristiques des « micros-unités » qui le composent (individus, foyers...). Cette modélisation se base sur la réplication des comportements de ces « micro-unités ». Ainsi, un modèle de microsimulation agit sur une base de données individuelles qui est vieillie de période en période. Les résultats pour chacune des unités étudiées sont ensuite agrégés.

La mise en œuvre de cette méthode suppose tout d'abord la constitution d'un fichier de données individuelles représentatif de la population que l'on souhaite modéliser. Ensuite les changements des caractéristiques des entités sont définis en effectuant des hypothèses d'évolutions. Chaque changement de caractéristiques est défini par des probabilités de passage d'un état A vers un état B en prenant en compte les relations entre les différentes variables. Ces paramètres sont obtenus par analyses statistiques de l'évolution antérieure de la population à projeter.

La simulation des évolutions individuelles se fait de façon déterministe pour certains événements (par exemple l'incrémentación de l'âge à chaque pas de la simulation). Pour d'autres événements, les évolutions se font de manière stochastique et repose sur l'utilisation de probabilités : la simulation a recours à la méthode de Monte Carlo qui consiste en une succession de tirages pseudos aléatoires.

A titre d'illustration, considérant l'événement décès pour un individu d'âge i , ce dernier a une probabilité de décéder dans l'année égale à q_i qui dépend de son âge et de son sexe. Soit u_i un nombre aléatoire tiré dans une loi uniforme $[0;1]$. Alors en comparant u_i et q_i , il est décidé si l'individu simulé décède ou non : si $u_i < q_i$ la mort de l'entité est simulée et dans le cas contraire, l'entité continue de vivre virtuellement.

La microsimulation s'appuie donc sur la loi des grands nombres car implicitement, le nombre de répétitions de l'expérience aléatoire dans l'échantillon considéré est si grand que le nombre projeté d'événements sera approximativement égal à son espérance.

En effet, selon la loi des grands nombres, si Y est le nombre d'individus connaissant l'événement simulé pour une population de taille N suffisant alors si q_i est la probabilité de survenance de l'événement, 95% des cas qui connaîtront l'événement sont dans l'intervalle :

$$N \times q_i \pm 1.96 \times \sqrt{q_i \times (1 - q_i) \times N}$$

Si par exemple l'événement simulé est le décès avec probabilité $q_i = 5\%$, $N = 10\,000$ le nombre Y de décès dans l'année sera dans 95% des cas dans l'intervalle $[478;522]$.

Les principes fondamentaux de la microsimulation sont les suivants :

- la microsimulation considère un échantillon de population plutôt qu'une population totale.
- la microsimulation nous fait travailler au niveau individuel.
- la microsimulation se base sur des tirages aléatoires plutôt que sur des proportions moyennes.

Dans le but de présenter la microsimulation nous nous attacherons tout d'abord à justifier le choix de cette méthode par analyse de la méthode des composantes appliquée à notre sujet. Dans la seconde partie, nous discuterons des points forts et des points faibles de la microsimulation. Enfin la dernière partie présentera un panorama des différents principaux modèles de microsimulation utilisés.

A. La méthode des composantes

1. Principe

Comme nous l'avons énoncé précédemment, historiquement, la méthode des composantes était la plus utilisée pour les projections démographiques. Nous allons présenter ici ses principaux mécanismes dans un cadre général d'évolution de population.

La base fondamentale de cette méthode est d'estimer, année après année, le nombre de naissances, de décès ainsi que le solde migratoire. Le solde migratoire est la différence entre le nombre de personnes qui sont entrées sur le territoire et le nombre de personnes qui en sont sorties au cours de l'année. Ce concept est indépendant de la nationalité.

Le point de départ de la projection est la population au 1er janvier de l'année N. Alors la population projetée au 1er janvier N+1 est :

$$Population_{N+1} = Population_N + Naissance_N - Décès_N + Solde migratoire_N$$

Les hypothèses de projections portent sur les taux de fécondité par âge de l'année, les quotients de décès par sexe et âge de l'année et le solde migratoire par sexe et âge de l'année.

Pour les naissances :

Chaque année, le nombre de naissance est obtenu en appliquant à chaque femme en état de procréer le taux de fécondité projeté de l'année considérée (ces taux dépendent de l'âge). En général, il est admis que la population féminine féconde est celle des 15-45 ans. Alors en considérant qu' a est l'âge des femmes atteint au 1^{er} janvier N+1, le nombre de naissance vaut :

$$Naissances_N = \sum_{a=15}^{45} (Population_{femme,N}^{a-1} + \frac{Solde migratoire_{femme,N}^a - Décès_{femme,N}^a}{2}) \times Taux\ fécondité_N^a$$

A la population de femmes d'âge $a-1$ au 1^{er} janvier N il est ajouté la différence entre le solde migratoire et les décès des femmes de même âge divisé par deux pour tenir compte des variations d'effectifs en cours d'année. Par convention, on estime que la moitié des femmes qui émigre et la moitié des femmes qui décèdent dans l'année sont soumises au « risque » de naissance.

En général pour déterminer le sexe de l'enfant, on considère qu'il y a sensiblement plus de naissances de garçons que de filles (Environ 105 garçons pour 100 filles en Amérique et Europe).

Pour les décès :

Le nombre de décès d'hommes ou de femmes d'âge a est obtenu en appliquant un taux de mortalité par âge et par sexe à la population moyenne de l'année N. Cette population moyenne est obtenue en ajoutant à la population d'âge $a-1$ au 1^{er} janvier N la moitié des entrées nettes d'hommes et de femmes d'âge a . Si s est le sexe d'une personne de la population, alors le nombre de décès au cours de l'année n est obtenu selon la formule :

$$Décès_{N,s}^a = (Population_{N,s}^{a-1} + \frac{Solde\ migratoire_{N,s}^a}{2}) \times Taux\ mortalité_{N,s}^a \text{ si } a > 0$$

Le nombre de décès des nouveaux nés est calculé de la façon suivante :

$$Décès_{N,s}^0 = Naissances_{N,s} \times Taux\ mortalité_{N,s}^0$$

La population au 1^{er} janvier N+1 se déduit donc de ce qui précède :

$$Population_{N+1,s}^a = Population_{N,s}^{a-1} - Décès_{N,s}^a + Solde\ migratoire_{N,s}^a \text{ si } a > 0$$

$$Population_{N+1,s}^0 = Naissances_{n,s} - Décès_{N,s}^0 + Solde\ migratoire_{N,s}^0 \text{ sinon}$$

Alors en sommant par âge et par sexe on obtient la population totale au 1^{er} janvier N+1.

Il est commun pour ces projections de population d'avoir recours à différents scénarios pour chacune de ces trois composantes : une hypothèse haute, une hypothèse basse et une hypothèse centrale. Il est alors possible de composer plusieurs scénarios en assemblant les différents niveaux d'hypothèse disponibles. On obtient donc 27 scénarios distincts (3x3x3).

A titre d'illustration, ci après les 27 scénarios produits par l'INSEE lors de l'étude de projection de la population française à horizon 2005-2050 :

Nom du scénario	fécondité	Espérance de vie	migration
Central	Centrale	Centrale	Centrale
1^{er} groupe de variantes : les 6 scénarios qui ne diffèrent du scénario central que par une seule composante			
Fécondité haute	Haute	centrale	Centrale
Fécondité basse	Basse	Centrale	Centrale
Espérance de vie haute	Centrale	Haute	Centrale
Espérance de vie basse	Centrale	Basse	Centrale
Migrations hautes	Centrale	Centrale	Haute
Migrations basses	centrale	centrale	basse
2^{ème} groupe de variantes : les 2 scénarios qui conduisent en 2050 au plus grand (plus petit) nombre d'habitants ; et les 2 scénarios qui conduisent à la population la plus jeune et la plus âgée			
Population haute	Haute	Haute	Haute
Population basse	Basse	Basse	Basse
Population jeune	Haute	Basse	Haute
Population âgée	Basse	Haute	Basse
3^{ème} groupe : autres variantes			
<i>Avec hypothèse centrale de fécondité</i>			
Espérance de vie basse-migrations basses	Centrale	Basse	Basse
Espérance de vie haute-migrations hautes	Centrale	Haute	Haute
Espérance de vie haute-migrations basses	Centrale	Haute	Basse
Espérance de vie basse-migrations hautes	Centrale	Basse	Haute
<i>Avec hypothèse centrale de migrations</i>			
Fécondité haute-espérance de vie haute	Haute	Haute	Centrale
Fécondité haute-espérance de vie basse	Haute	Basse	Centrale
Fécondité basse-espérance de vie haute	Basse	Haute	Centrale
Fécondité basse-espérance de vie basse	Basse	Basse	Centrale
<i>Avec hypothèse centrale de mortalité</i>			
Fécondité basse-migrations hautes	Basse	Centrale	Haute
Fécondité basse-migrations basses	Basse	Centrale	Basse
Fécondité haute-migrations hautes	Haute	Centrale	Haute
Fécondité haute-migrations basses	Haute	Centrale	Basse
<i>Sans hypothèses centrales</i>			
Fécondité et espérance de vie hautes, migrations basses	Haute	Haute	Basse
Fécondité haute, espérance de vie et migrations basses	Haute	Basse	Basse
Fécondité et migrations basses, espérance de vie haute	Basse	Haute	Basse
Fécondité, migrations et espérance de vie basses	Basse	Basse	Basse

Source : Insee, projections de population 2005-2050

2. La méthode des composantes dans notre étude

Nous avons vu que les projections déterministes partent d'une description de la population selon un nombre limité de paramètres et cette description peut donc se subdiviser en classes sous forme d'un vecteur d'état.

Le passage d'un état entre t et $t+1$ peut alors se résumer à l'application d'une matrice de passage A dont l'élément a_{ij} équivaudra en général à la probabilité pour un individu de la catégorie j de se retrouver en $t+1$ dans la catégorie i . On a alors :

$$P_{t+1} = A_t P_t \text{ et } P_{t+n} = A_{t+n-1} A_{t+n-2} \dots P_t$$

Comme nous l'avons dit, le cas particulier usuel est la structure par sexe et âge d'une population. Alors en ne considérant que les femmes, si on note $p_{a,t}$ l'effectif d'âge a à l'instant t , $q_{a,t}$ le quotient de mortalité à l'âge a et $f_{a,t}$ la probabilité de donner naissance à une fille à cette âge a on voit que la population à l'instant $t+1$ rangée sous forme de vecteur colonne peut s'écrire :

$$P_{t+1} = \begin{bmatrix} p_{0,t+1} \\ \vdots \\ p_{15,t+1} \\ \vdots \\ p_{45,t+1} \\ \vdots \\ p_{99,t+1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & f_{15,t} & \dots & f_{45,t} & 0 & 0 \\ 1-q_{0,t} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1-q_{1,t} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1-q_{15,t} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-q_{45,t} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1-q_{98,t} & 0 \end{bmatrix} \begin{bmatrix} p_{0,t} \\ \vdots \\ p_{15,t} \\ \vdots \\ p_{45,t} \\ \vdots \\ p_{99,t} \end{bmatrix}_t$$

$$= \begin{bmatrix} \sum_{k=15}^{45} p_{k,t} \times f_{k,t} \\ \vdots \\ (1-q_{15,t}) \times p_{15,t} \\ \vdots \\ (1-q_{45,t}) \times p_{45,t} \\ \vdots \\ (1-q_{98,t}) \times p_{98,t} \end{bmatrix} = AP_t$$

Le passage à une population découpée selon des critères plus nombreux que le sexe et l'âge se fait ensuite en accroissant la dimension du vecteur P est en complexifiant la matrice A.

Cette méthode se heurte donc à plusieurs limites :

- le partage en classes ne vaut que pour des variables discrètes, par exemple pour projeter une population par niveau de revenus il faudra procéder à des regroupements par classes de revenu ce qui induit une perte d'information non négligeable car cela forcerait l'utilisation de probabilité de transition.
- la dimension de la matrice à gérer croît de façon exponentielle avec le nombre de paramètres pris en compte.
- ne peuvent être projetés par le modèle que les indicateurs calculables à partir du vecteur P. Le découpage en classes prévu au départ conditionne donc l'ensemble des résultats disponibles. L'obtention de résultats supplémentaires suppose alors d'affiner le découpage et augmente en conséquence la taille du vecteur P.

Nous pouvons donc dire que cette méthode, appliquée à la projection de population, se concentre essentiellement sur des proportions moyennes : l'individu, avec toutes ses spécificités ne peut trouver sa place dans cette modélisation, seul le groupe est pris en considération pour la détermination des hypothèses d'évolution. Cela a pour conséquence de limiter le nombre de paramètres pouvant être liés avec les autres pour prendre en compte l'ensemble des comportements.

La microsimulation, en raisonnant au niveau des individus, permet de prendre en compte un plus grand éventail de paramètres modélisant les comportements. Nous allons donc étudier dans quelle mesure cette méthode est particulièrement adaptée dans le cadre de ce mémoire dans la section suivante.

B. Caractéristiques de la microsimulation

1. Les différents modèles de microsimulation

Il existe deux principaux modèles de microsimulation : les modèles statiques et les modèles dynamiques.

Les premiers sont souvent utilisés pour mesurer l'impact des politiques fiscales ou sociales sur une année. La structure des « micro-unités » de base dans l'échantillon n'est pas amenée à varier même si les entités peuvent être vieillies. Les individus d'un groupe ne sont pas supposés changer de comportement avec le temps. Ces modèles sont donc surtout utilisés pour des horizons court ou moyen terme (de 1 à 5 ans).

Les modèles dynamiques sont construits à partir d'observations transversales ou longitudinales. Les individus sont suivis pendant leur existence en simulant différents événements (naissance, mort, revenus, pensions de retraite, santé...). Les données de l'échantillon sont actualisées à chaque période : il s'agit de « *mettre à jour chaque caractéristique pour chaque micro-unité pour chaque intervalle de temps* » (Caldwell, 1990, p.5). De ce fait, les individus peuvent être amenés à changer de comportements en fonction de leur évolution ; par exemple, une femme mariée depuis peu n'a pas le même comportement face à la natalité qu'une femme célibataire ayant les mêmes caractéristiques. Ce type de modèle est particulièrement apprécié lors de simulations des politiques publiques à long terme.

Ces modèles dynamiques se subdivisent en deux catégories :

- Les modèles longitudinaux qui simulent l'ensemble de la vie d'une seule génération (de la naissance à la mort). Dans ce type de modèle, la vie d'un individu est entièrement simulée avant de passer à l'individu suivant. Ceci pose le problème majeur d'absence d'interaction entre chaque micro-unité.

- Les modèles transversaux pour lesquels toutes les unités d'un échantillon de la population sont simulées sur une période de temps donnée. De cette façon, les interactions sont possibles entre les individus.

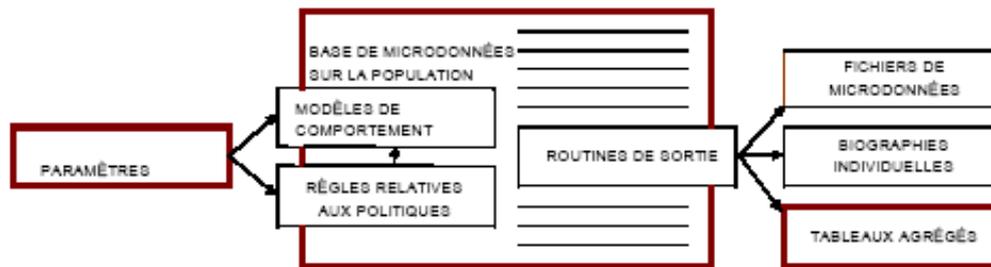
Parmi tous ces modèles il faut distinguer ceux qui sont ouverts ou fermés, en temps continu ou en temps discret, comportementaux ou probabilistes.

Les modèles probabilistes tentent de reproduire une structure observée au sein d'un échantillon. La modélisation des événements n'a aucun fondement théorique et se base donc uniquement sur l'observation empirique. Contrairement à ce dernier, le modèle comportemental se base sur des fondements théoriques. Cependant, ces modèles sont peu nombreux et leur usage dans des modèles de microsimulation s'avère délicat et augmente de façon significative le temps d'exécution d'une simulation. De plus, Pudney et Sutherland (1996) ont montré que les estimations fondées sur ces modèles de comportement étaient soumises à des intervalles de confiance plus importants.

Les modèles en temps discret posent problème lorsque des transitions d'une durée inférieure au pas s'opèrent. Par exemple, avec un pas annuel, un individu dans la vie réelle peut passer par plusieurs états pendant ce laps de temps. Il semble alors préférable de s'orienter vers des modèles en temps continu qui se basent sur l'analyse de survie. Selon ce que l'on souhaite modéliser, par souci de simplicité et selon les événements entrant en jeu il est alors préféré l'utilisation des modèles à pas discret qui sont plus simple à mettre en œuvre.

Un modèle est dit fermé s'il s'appuie sur un nombre fixé d'individus qui ne varie que par le nombre de naissances et d'immigrants, les mariages se font entre les entités déjà existantes. A l'opposé, un modèle est qualifié d'ouvert lorsque des entités sont créés ex-nihilo¹ lors des mariages. En pratique il est toujours plus facile de travailler avec des modèles fermés car il est plus aisé de faire des ajustements sur les agrégats si le poids des différents individus reste constant à chaque période.

¹ De nouveaux individus sont créés virtuellement avec des caractéristiques proches des micro-unités en relation.



Composantes d'un modèle de microsimulation. Source : Statistiques Canada

Les caractéristiques de la microsimulation sont généralement mises en relief par rapport à la macrosimulation et ont été développées, entre autre, dans l'article de la revue « Population » numéro 4 de 1997 par E. VAN IMHOFF et W.POST.

2. Atouts et limites de la microsimulation

En considérant de nouveau l'exemple précédemment décrit, on peut affirmer que la macrosimulation et la microsimulation s'appuient sur la loi des grands nombres : si l'un considère que la taille de population est si grande que le nombre de décès projetés peut être posé comme égal à son espérance mathématique, l'autre suppose que le nombre de répétitions du tirage aléatoire est suffisant pour que le nombre projeté de décès soit approximativement égal à son espérance.

De ce fait, la microsimulation est capable de nous donner une projection mais aussi un intervalle de confiance : en réitérant plusieurs fois la projection, nous obtiendrons plusieurs résultats différents, ce qui nous permettra de calculer directement l'écart type. Chose que la macrosimulation ne peut faire dans la plupart des cas : on obtiendra toujours la même projection, de plus, les calculs d'écart types sont compliqués et leur mise en œuvre s'avère très délicate.

Ces deux modèles n'ont pas la même approche concernant le traitement des données en vue de la spécification des équations sur les comportements : comme leur raisonnement ne se situe pas au même niveau, la façon dont sont étudiées les données originelles diffère. Ainsi les modèles en microsimulation souffrent d'un fort besoin en données exhaustives afin de justifier leur utilisation. Ce travail préliminaire pouvant être difficile à mettre en œuvre trouve sa justification dans l'un de ses points les plus forts : un agent moyen ne permet pas d'expliquer la diversité des comportements économiques. La microsimulation en se situant au niveau de chaque micro-unité simule le devenir de chaque individu. Cette méthode permet de tenir compte de l'hétérogénéité des agents et fournit un éventail de résultats beaucoup plus large. Les sorties ne sont pas qu'un nombre d'individus avec leur caractéristiques mais autant de biographies individuelles : tous les scénarios en fin de projection peuvent être assemblés de façon presque illimitée.

Une autre différence entre ces deux modèles concerne l'implémentation informatique. Si pour les modèles macro il existe pléthore de logiciels performants pouvant s'adapter à beaucoup de cas, chaque programme micro est unique et ne peut se transposer à une autre étude en

microsimulation. De plus leur élaboration est beaucoup plus difficile à mettre en œuvre. Les premiers modèles ont été conçus en plusieurs années.

Dans l'optique de comparaison des prises en compte des effets d'interactions entre variables, on peut affirmer que le choix de la microsimulation est plus opportun. En effet, pour tous les comportements possibles et les équations les régissant, la probabilité d'un événement au niveau individuel pourra être déterminée immédiatement car les variables explicatives sont disponibles. Or, dans un modèle macro, il faudra, pour les mêmes considérations d'interactions entre les variables, considérer un très grand espace d'états qui est problématique au niveau macroéconomique comme expliqué auparavant.

Un autre avantage est que les liens régissant les relations entre les individus sont clairement définis en microsimulation. Typiquement, les mariages ou les divorces font intervenir plusieurs personnes et en microsimulation nous pouvons décrire les liens d'un individu avec l'ensemble de la population alors qu'en macrosimulation, ceci n'est pas possible de part les fondements de sa modélisation : il existe une « mémoire » des liens tissés par un individu au cours de son évolution projetée en microsimulation. Ceci pose tout de même le problème d'appariement : chaque nouvel individu marié doit être relié à un conjoint existant par exemple.

Pour information, seule la microsimulation peut faire intervenir des variables continues. Elle peut notamment faire intervenir des modèles de durée afin de tenir compte du caractère continu des événements.

Sur la base de ces considérations, nous pouvons conclure que la microsimulation est plus à même de répondre à nos besoins. Elle est un outil puissant dès lors qu'il s'agit de travailler avec un grand nombre de variables interdépendantes ou lorsqu'elles sont continues. Cependant, pour estimer les paramètres nécessaires aux équations comportementales, l'étude préliminaire doit être fine et peut être complexe à mettre en œuvre.

De par leur nature, les modèles ne pourront jamais donner un résultat certain dans leurs prédictions. En effet, même avec la modélisation la plus fine possible, il sera toujours illusoire de connaître avec certitude les comportements humains. On pourra dans la plupart des cas

donner une « tendance » pour un individu ayant certaines caractéristiques à préférer un choix plutôt qu'un autre mais cela ne pourra jamais coïncider totalement avec la réalité.

Ceci est valable pour tous les modèles de simulation des comportements humains, qu'ils soient au niveau macro ou micro. Cependant, la microsimulation connaît une dépendance plus forte liée à l'aléa.

3. L'aléa en microsimulation

L'aléa occupe une place prépondérante dans la modélisation par microsimulation. Comme nous l'avons expliqué précédemment, le principe est de soumettre chaque individu de l'échantillon à l'aléa pour le faire évoluer sur chaque intervalle de temps. De ce fait, en relançant plusieurs fois la simulation pour un échantillon identique et même pour un nombre de « choix » limités par individu, la simulation ne donnera pas toujours les mêmes résultats. Un des moyens pour minimiser ces écarts est d'augmenter la taille de l'échantillon considéré. S'ajoute à cela le choix du générateur de la loi uniforme qui doit avoir certaines spécificités. Bon nombre de générateurs implémentés dans les logiciels usuels ne doivent pas être utilisés car obsolètes.

Le choix de l'échantillon de base est une source d'aléa. En effet, même en considérant que l'aléa inhérent est extrêmement faible, il reste à définir un échantillon qui soit fidèle à la population de départ et qui de part son évolution, puisse définir de la manière la plus exacte possible celle de la population totale. Un changement de l'échantillon de base peut amener la projection de la population à varier. Encore une fois, un des moyens pour parer cet aléa reste l'augmentation de la taille de l'échantillon cible, et pour l'éliminer totalement il faudrait considérer la population initiale dans sa globalité. Or, ceci n'est possible que pour un nombre de cas limités ou la population considérée est de taille raisonnable.

Ces deux aléas sont donc sources de variations aléatoires de la projection mais peuvent être maîtrisés et minimisés en augmentant la taille de l'échantillon sans pour autant pouvoir être éliminés sauf cas exceptionnel.

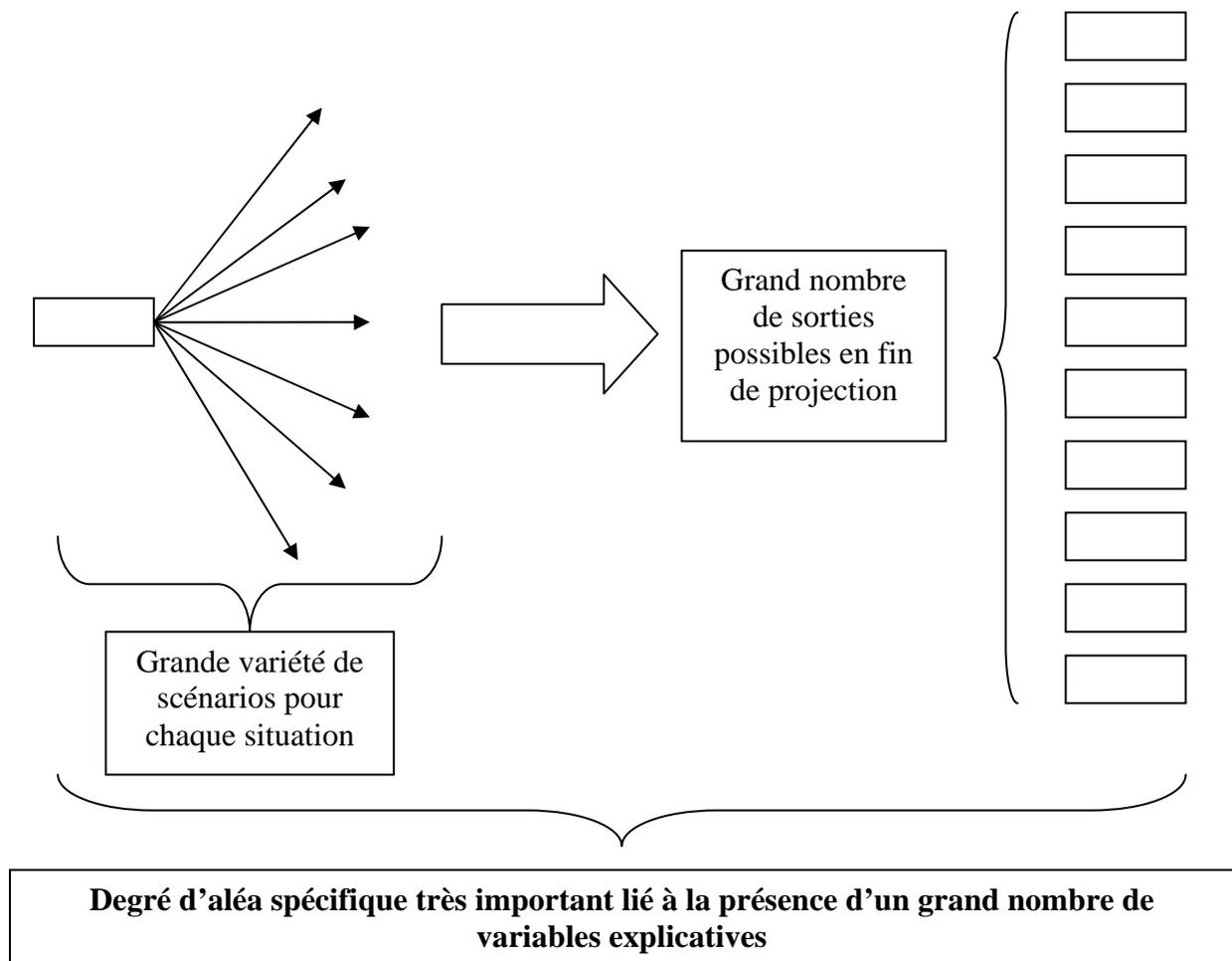
Pour autant, ces deux aléas sont négligeables en comparaison à l'aléa de spécification² comme nous allons le montrer.

Afin de projeter un échantillon donné, il faut en premier lieu procéder à une étude économétrique de l'évolution antérieure de la population considérée. Il faut donc dans le même temps être assez fin dans cette analyse et lier autant de variables que possibles afin de tenir compte de la complexité des comportements sans pour autant tomber dans le cas extrême

² Pudney et Sutherland, 1994

où il y a dépendance totale des variables. Ceci pose le problème de spécification : à trop définir de variables explicatives, on augmente le degré d'aléa des sorties. En effet, si on modélise trop de scénarios probables dans une même situation, une sortie pourra prendre autant de valeurs différentes. Les caractéristiques de fin de projections d'un individu pourraient donc avoir une multitude de parcours différents :

L'aléa de spécification



Cependant, en diminuant trop le nombre de variables explicatives, il est alors évident que le risque est de ne pas prendre en compte la spécificité de la population et la microsimulation n'est alors plus capable de tirer profit de ses atouts et donc son utilisation ne se justifie plus.

Le problème est donc de trouver un compromis entre l'arborescence des variables explicatives et leur limitation pour éviter respectivement un haut degré d'aléa de spécification et l'erreur de spécification.

4. La hiérarchisation des risques

En temps discret avec un pas annuel, il peut survenir plusieurs événements pour un même individu : mariage puis naissance d'un enfant puis divorce par exemple. Ceci pose le problème des événements multiples : quelle priorité donner aux événements les uns par rapport aux autres ? Par exemple, faut-il d'abord simuler les décès de l'échantillon avant de considérer les naissances ?

La façon dont seront hiérarchisés ces risques influera le devenir de la projection de la population : si on donne systématiquement priorité à la mort d'une femme avant de donner naissance à un enfant, cela aura un impact en fin de projection. Alors en choisissant arbitrairement pour tous les risques plusieurs hiérarchisations on peut avoir des résultats globaux de projections différents. Il faut donc hiérarchiser les événements simulés tout en restant conscient que cela a un impact sur les projections faites par la microsimulation.

5. Les dernières avancées en microsimulation

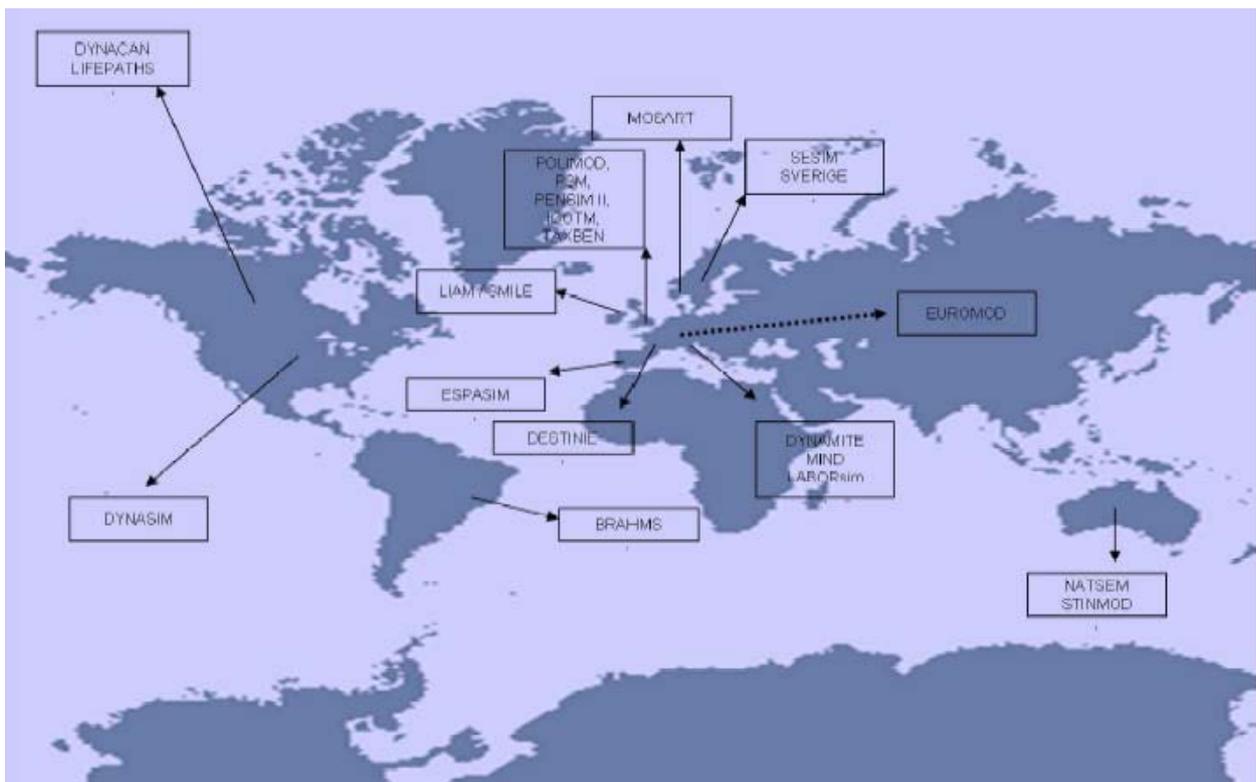
Historiquement, on attribue la paternité de la microsimulation à Guy Orcutt en 1957 lors de la parution de son article « *A new type of socio economic system* » dans *Review of Economics and Statistics*. Les moyens techniques de l'époque n'étaient pas aussi avancés qu'aujourd'hui et les applications en microsimulation n'ont été développées que très récemment. Nous nous proposons ici de citer les principaux modèles de microsimulation existants à ce jour.

Le modèle américain, DYNASIM, parmi les premiers à avoir été créés, comporte trois versions dont la première bâtie dans les années 70 a pris six ans de conception. La dernière version en date se base sur un échantillon du *Survey of Income and Program Participation Panel* de 1990-1993. Ce nouveau modèle comprend plusieurs modules d'épargne et de pension de retraite privée pour un horizon de projection 2030-2040.

LIFEPATHS est le modèle de microsimulation du Canada développé par Statistiques Canada. C'est un modèle longitudinal dynamique de microsimulation des individus et des familles. À l'aide d'équations sur le comportement fondées sur des données chronologiques, il crée des échantillons statistiquement représentatifs de cheminements de vie individuels complets. Il est utilisé pour analyser et élaborer des politiques gouvernementales ayant une composante essentiellement longitudinale, particulièrement celles qui, de par leur nature, exigent une évaluation au niveau de l'individu ou de la famille. Il peut aussi être utilisé pour analyser diverses questions sociales de nature longitudinale comme l'équité entre les générations et l'emploi du temps tout au long de la vie.

Par ailleurs, Statistiques Canada a aussi développé Modgen. C'est un outil général créé pour faciliter la construction et l'utilisation de modèles longitudinaux de microsimulation. La structure, les équations, la documentation et les tableaux de résultats d'un modèle sont entrés dans Modgen au moyen d'un langage de haut niveau. Modgen construit ensuite un programme informatique entièrement fonctionnel doté d'une interface utilisateur qui met en œuvre le modèle spécifié. BioBrowser, une application associée à Modgen, permet d'explorer visuellement les profils longitudinaux des entités simulées de bas niveau générés par les modèles Modgen.

Le modèle français, DESTINIE, a été créé en 1999. Ce modèle simule le vieillissement et le renouvellement d'un échantillon d'environ 15 000 ménages et 37 000 individus issu de l'enquête Actifs Financiers 1991 de l'INSEE, par une combinaison de règles déterministes, de tirages aléatoires et de modèles comportementaux. Le modèle étant principalement construit pour analyser l'évolution de la situation des retraités, une attention particulière a été portée à la modélisation des événements démographiques (naissances, décès, unions, séparations) et aux carrières salariales (niveau de salaire, passages éventuels par le chômage ou l'inactivité). Dans sa version actuelle, le modèle considère que tous les individus font partie du secteur privé et perçoivent à leur retraite des pensions du régime général et des régimes complémentaires obligatoires ARRCO et AGIRC. Le modèle permet ainsi de simuler les retraites jusqu'à l'horizon 2040 et d'analyser les effets redistributifs (inter et intragénérationnels) d'une modification des règles des régimes de retraite ou plus généralement de phénomènes structurels marquants (croissance de l'espérance de vie, montée de l'activité féminine, allongement de la durée des études).



Un panorama non exhaustif des modèles de microsimulation. Labor [2007]

Beaucoup de ces modèles intègrent des modules comportementaux et plus particulièrement ceux ayant trait à la retraite. En effet, les changements démographiques étant mondiaux, toutes les nations sont concernées par des changements législatifs des régimes de retraite et tentent donc de mesurer l'impact de ces changements.

Le modèle que nous avons construit est bien évidemment une version très simplifiée par rapport aux modèles nationaux existants : seuls les évènements démographiques (naissances, décès) et les paramètres ayant une influence sur les mariages et les divorces en général (durée+occurrence) ont été considérés.

La partie suivante décrit l'analyse des données qui ont été en notre possession ainsi que la construction de la base en entrée du modèle. Nous nous attacherons ensuite à expliciter le calibrage des paramètres de projection. Enfin nous détaillerons la programmation de la première version du modèle ainsi que les premiers résultats.

Partie II : Travaux réalisés et 1ers résultats.

« *Au mieux les modèles sont aussi bons que les données sur lesquels ils reposent* » (Harding, 1993)

A. La construction de la base de données initiale

Pour réaliser les projections, il fallait que nous puissions disposer d'une base de données initiale sur la population, avec suffisamment d'information les concernant.

Les données qui nous ont été fournies :

Nous avons disposé des informations suivantes :

- Evolution de la population cible entre 1984 et fin 2009 (nombre de décès, mariages, divorces, naissances...),
- Statistiques sur le nombre de divorces de couples par durée de mariage,
- Détail de la composition des foyers au 31.12.2008.

Concernant le détail de la composition de chaque foyer au 31.12.2008, les données à disposition pour chaque individu sont :

- Numéro de ménage,
- Numéro d'individu dans le ménage,
- Sexe,
- Bénéficiaire ou pas de l'avantage,
- Situation familiale (célibataire, marié(e), veuf (ve), divorcé(e)),
- Année de naissance.

Ce fichier nous a servi de base à la projection de la population. Cependant, il n'était pas directement utilisable, et a dû faire l'objet d'un retraitement comme expliqué ci après.

1. Les retraitements effectués

Le fichier du détail de la composition des foyers comportait les données individuelles suivantes :

- Numéro de ménage.
- Numéro d'individu dans le ménage.
- Sexe.
- Bénéficiaire ou pas de l'avantage.
- Situation familiale (Célibataire, marié(e), veuf (ve), Divorcé(e)).
- Année de naissance.

A partir de ce fichier, nous n'avons gardé que la population cible, en reconstituant pour chacun d'eux un lien familial avec les autres membres du même foyer, afin d'identifier pour les individus mariés, leurs ayants droits.

La répartition de la population est la suivante :

- Foyers à un individu : Toutes les situations familiales sont représentées. Ne comprenant qu'un seul individu, ces ménages n'ont pas fait l'objet d'un appariement spécifique. A noter la présence de mariés (tous issus de la population cible) dont le conjoint n'est pas répertorié. Nous avons attribué à ces personnes un conjoint qui n'a pas encore demandé l'avantage.
- Foyers à deux individus : Ici les liens d'appariement étaient facilement identifiables grâce à la situation familiale, couplée à l'année de naissance : parent-enfant, mari-femme, personnes en concubinage.
- Foyers de 3 à 5 individus : On peut penser qu'il s'agit pour la très grande majorité de familles comportant une, deux voir trois générations. Encore une fois, nous avons apparié les mariés dans un même foyer en tentant d'identifier les enfants probables (minimiser la différence d'année de naissance pour l'appariement des maris et des femmes a été un critère).
- Foyers de 6 à 10 individus. Même mode opératoire que précédemment.

- Foyers de plus de 18 individus : Il s'agit probablement de foyers regroupant plusieurs personnes n'ayant aucun lien de parenté entre eux hormis les personnes mariées...

Il nous manquait dans ce fichier une information importante pour la projection, qui est, pour les personnes mariées, la date du mariage, déterminant la durée du mariage et donc la possibilité pour un « étranger » de prétendre à l'avantage ou non.

Les données recueillies ne nous donnant aucune information sur cette variable, nous avons en conséquence travaillé avec les statistiques de l'INSEE (Institut National de la Statistique et des Etudes Economiques) concernant l'âge moyen au premier mariage et les probabilités de divorce et de remariage depuis cette date de premier mariage.

Pour chaque couple marié, nous avons, en fonction de l'âge de chacun des époux, déterminé un âge théorique de mariage, puis à l'aide d'une fonction d'arbitrage entre les deux dates, nous avons retenu une date théorique de mariage.

Concernant le très faible nombre de personnes dont la situation familiale n'était pas renseignée, nous avons considéré qu'elles étaient célibataires.

Nous avons enlevé de la base de données quelques individus dont l'année de naissance renseignée était 1900 (dû au fait que l'information n'était pas disponible), en effet ces personnes biaisaient la mortalité de façon significative.

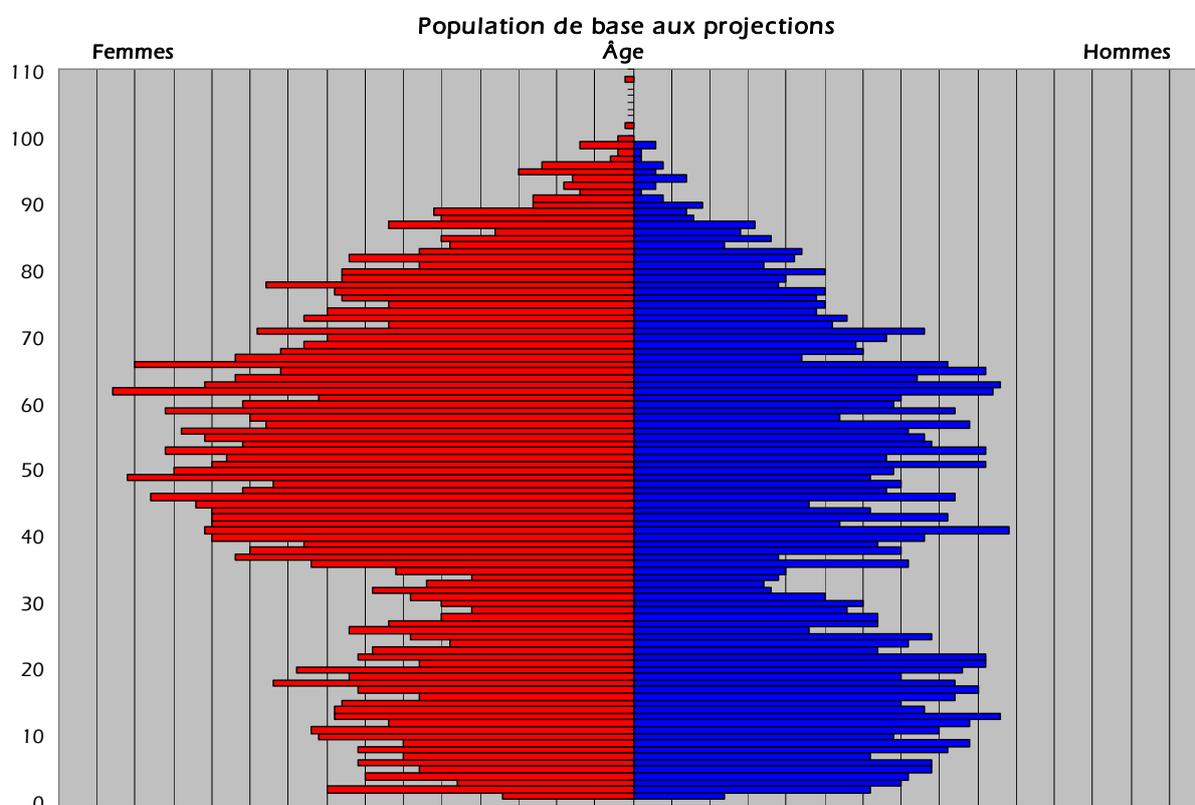
Après retraitement, nous avons obtenu une base de données. La description de la base des données retenues est donnée ci-après.

2. La description de la base de données finale

Nous avons donc en entrée des projections un fichier de données individuelles. Ce fichier comportait les informations suivantes :

- Numéro de ménage.
- Numéro d'individu dans le ménage.
- Numéro d'identification unique
- Sexe.
- Bénéficiaire ou pas de l'avantage.
- Situation familiale (célibataire, marié(e), veuf (ve), divorcé(e), en couple) avec identification des liens familiaux.
- Durée du mariage (théorique).
- Année de naissance.

La population de base des individus se répartie comme suit :



La pyramide des individus présente des « anomalies »:

- un creux démographique très marqué aux environs de 30 ans,
- à l'inverse, une bosse démographique importante chez les hommes d'environ 50 ans,
- une faible natalité (mais qui correspond à la natalité des classes creuses, en âge de procréer).

Un retraitement différent a été testé, notamment en diminuant l'importance du creux démographique constaté autour de 30 ans. Les résultats correspondants figurent en partie IV.A.

La base de données constituée considérée comme fiable a pu faire l'objet de tests dans son ensemble afin de fixer les paramètres de projections. La section suivante détaille le calibrage de ces hypothèses.

B. Le calibrage des paramètres de projections

Ce fichier final couvrait un peu plus de 80 % de la population totale.

Nous avons considéré qu'il était représentatif de la population globale, et avons en conséquence recalé l'ensemble de nos projections d'un facteur d'environ 20%.

Cette hypothèse pourrait être la source d'un léger biais dans nos projections, s'il s'avérait que les 20 % de la population cible non recensés ne présentent pas des caractéristiques moyennes identiques à celles de la population traitée.

1. La natalité

Nous avons appliqué les taux de fécondité de l'INSEE aux femmes, sans aucune déformation des tables. En ce qui concerne les hommes, nous avons différencié trois cas distincts :

- Les hommes mariés avec une femme issue de la population cible.
- Les hommes mariés avec une « étrangère ».
- Les hommes célibataires, veufs ou divorcés.

Pour les hommes mariés avec une femme issue de la population cible, le taux de fécondité a été supposé nul, considérant que cet événement est déjà simulé pour sa conjointe.

La fécondité des hommes mariés avec une « étrangère » a été construite de telle sorte qu'elle coïncide avec celle de leur conjointe : un décalage de 4 ans des taux a ainsi été opéré afin de tenir compte de la différence d'âge moyenne entre homme et femme au sein d'un couple marié.

Enfin, la fécondité des hommes non mariés a été construite de façon à prendre en compte la part croissante d'union libre : un décalage de 4 ans a aussi été opéré, couplé à une déformation des taux d'un facteur 0,5, afin de tenir compte du fait que sa partenaire est peut-

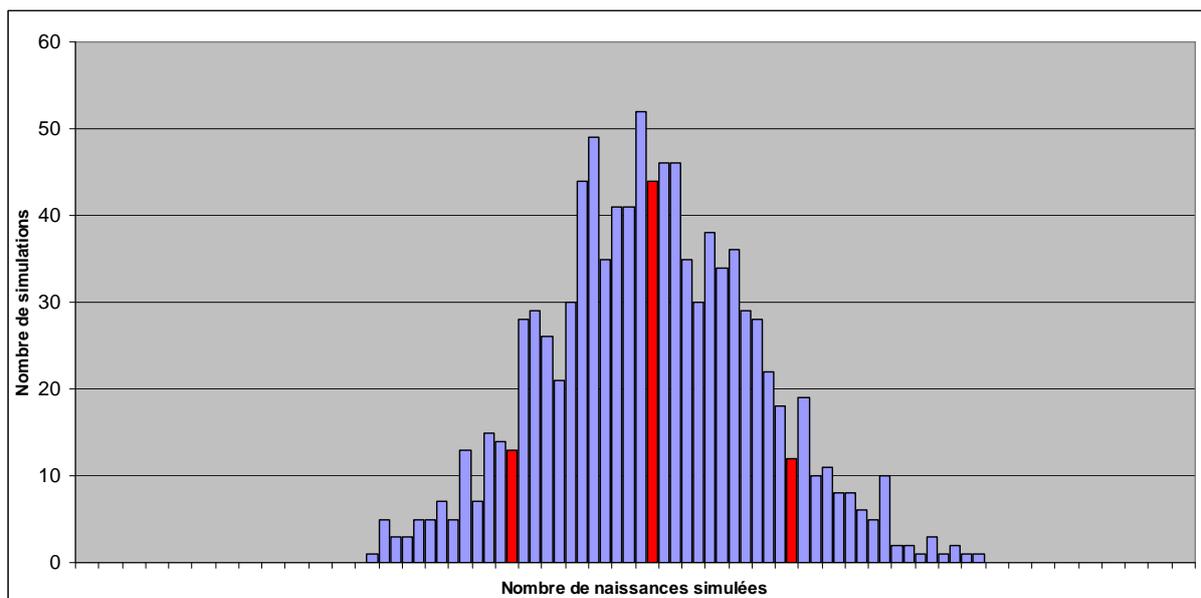
être une femme issue de la population de base, pour laquelle un taux de fécondité aurait déjà été simulé.

Le profil obtenu correspond à une fécondité simulée pour 80% de la population cible. Il faut en conséquence la ramener à l'ensemble des individus.

Selon l'historique des naissances par année entre 2000 et 2009, il existe un rapport de 0.80417 entre le nombre de naissances simulées avec les hypothèses précédentes pour l'ensemble de la population réelle théorique, et celui observé sur les dix dernières années.

Nous avons donc travaillé avec les taux décrits précédemment, retraités d'un coefficient 0.80417 pour les projections. En effet, il nous semble que les taux INSEE considérés sont particulièrement élevés du fait de la forte fécondité en France et retraiter ces taux de 20% semble en accord avec les caractéristiques de la population considérée, principalement citadine.

En effectuant 1 000 simulations et en négligeant les déciles inférieur et supérieur, représentés par un marqueur rouge dans le graphique ci-dessous (soit 800 simulations considérées) nous obtenons le profil de natalité suivant (La moyenne est aussi représentée par un marqueur rouge):



Les résultats semblent en accord avec l'historique des naissances depuis 2000, avec une moyenne égale à celle de l'historique sur 10 ans et avec des extrêmes également proches. Les taux de natalité INSEE sont donnés en annexe 2.

2. La mortalité

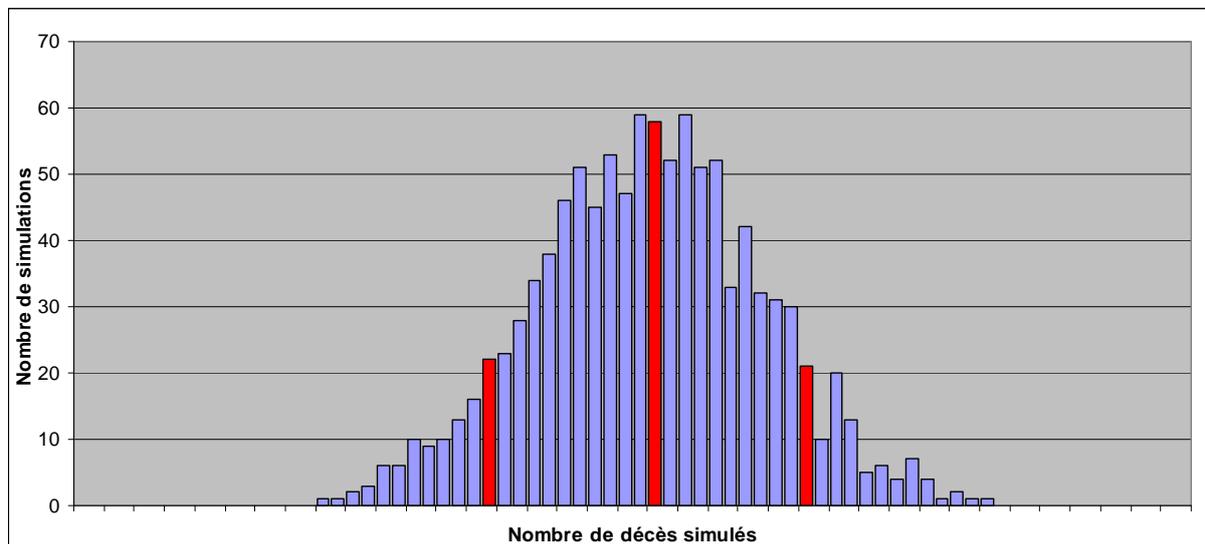
Nous avons appliqué des tables de mortalité générationnelles (Tables prospectives TGF-TGH 2005) pour 1 000 simulations. Comme pour la natalité, nous avons ramené le nombre de décès à l'ensemble de la population réelle.

En rapprochant l'historique de la mortalité avec les simulations effectuées, nous avons identifié un rapport de 0.9909 entre le nombre de décès observés et simulés pour l'ensemble de la population (retraitement des 20% effectué). La mortalité de la population étudiée est donc très légèrement inférieure à celle de l'ensemble de la population française.

Nous avons donc travaillé avec les tables de mortalité prospectives retraitées d'un coefficient 0.9909.

Les résultats semblent en accord avec l'historique de la mortalité tant au niveau de la moyenne qu'au niveau de l'étalement.

Nous obtenons le profil de mortalité suivant (1^{er} et 9^{ème} décile en rouge ainsi que la moyenne) :



3. La nuptialité

Nous avons appliqué à la population de base les tables INSEE des taux de nuptialité par sexe et groupe d'âge. Or, en France, nous disposons du PACS (Pacte civil de solidarité) depuis fin 1999 et à fin 2008, il a été observé un PACS de deux personnes de sexe opposé pour deux mariages.

La population étudiée ne disposant pas d'union officielle hors mariage, nous avons donc modifié les tables françaises de nuptialité en majorant de 50% les taux de mariage jusqu'à 39 ans. Il nous semble en effet que les PACS concernent principalement les individus les plus jeunes et le coefficient majorant s'explique par le rapport du nombre de PACS en France par rapport au nombre de mariages célébrés à fin 2008.

Comme précédemment, nous avons ramené le résultat des simulations à l'ensemble de la population réelle et avons observé un rapport de 1.097 entre le nombre de mariages d'individus de la population cible observé et simulé pour l'ensemble de la population.

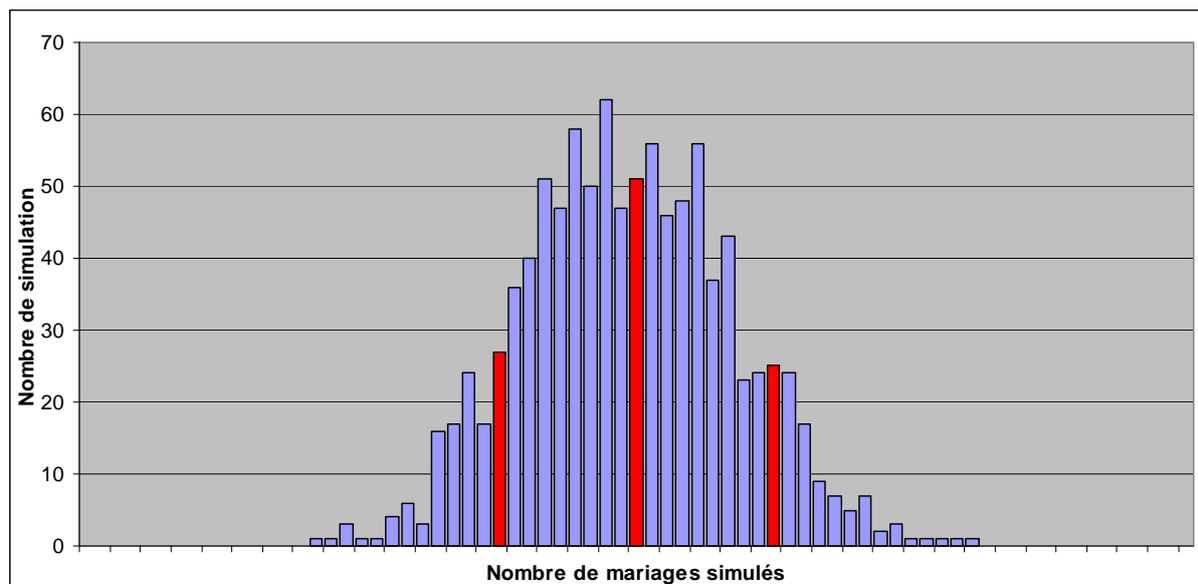
Il semble que la population cible se marie un peu plus que ce que nous le prévoyions avec les hypothèses précédentes. Nous avons en conséquence majoré les tables construites, d'un coefficient 1,097.

Une fois le mariage simulé, il s'agissait de l'origine du conjoint. En moyenne, depuis 1990, 5.48% des mariages concernent deux personnes issues de la population de base. Si l'on considère l'origine du conjoint de chaque individu de la population cible qui se marie, on arrive aux statistiques suivantes :

- 8,7 % des femmes se marient avec un homme issu de la population cible au moment du mariage,
- 8,4 % des hommes se marient avec une femme issue de la population cible au moment du mariage.

Nous avons donc considéré que 91,3% des mariages de femmes et 91,6 % des mariages d'hommes se font avec un « étranger », le reste des mariages se faisant entre individus issus de la population cible.

Les résultats semblent en accord avec l'historique des mariages des individus depuis 2000 et nous obtenons le profil suivant de nuptialité (1^{er} et 9^{ème} décile en rouge ainsi que la moyenne):



4. La divortialité

La question des divorces est différente des autres événements démographiques. En effet, l'avantage n'étant pas négligeable, il a toujours été observé une atténuation des divorces durant les 5 premières années de mariage. Ces derniers semblent s'être reportés dans les 3 ans qui suivent les 5 ans de mariages. Ce phénomène étant spécifique à la population étudiée, nous ne disposons pas de lois adaptées. Nous avons donc « transformé » les tables de divortialité de l'INSEE.

Afin de prendre en compte la réglementation actuelle concernant l'acquisition de l'avantage après 5 ans de mariage, nous avons opéré une déformation de la table INSEE des taux de divorce selon la durée du mariage : les taux de divorce pour les durées de mariage inférieures ou égales à 5 ans ont été diminués de 33%, et les taux de divorces entre 6 et 8 ans ont été majorés de 50%.

Le même mode opératoire a été appliqué concernant la réglementation envisagée : les taux de divorce pour les durées de mariage inférieures ou égales à 10 ans ont été diminués de 33%, et les taux de divorces entre 11 et 16 ans ont été majorés de 50%.

Nous avons donc « déplacé » une partie des divorces pendant les premières années et les avons reportés dans le futur. Ces deux tables donnent exactement le même taux de divorce pour un couple pour un horizon de 30 ans. Ce taux « long terme » est aussi égal à celui observé dans les statistiques INSEE. En effet, les données en notre possession ne permettent pas d'apprécier ce paramètre et nous avons calé les divorces sur les taux INSEE.

5. Les entrées

Nous avons introduit, en plus des lois précédentes, une loi d'entrée qui génère chaque année un nombre aléatoire de nouveaux individus.

Il a été demandé de ne modéliser que la moitié des entrées annuelles observées jusqu'ici en scénario central afin de tenir compte de l'atténuation probable de l'événement.

Chaque année et pour chaque individu parcouru pendant la simulation, nous avons effectué un tirage aléatoire et nous l'avons comparé au rapport nombre d'entrées voulues/Nombre d'individus total. Si le tirage aléatoire était inférieur à ce rapport alors un nouvel individu était créé possédant exactement les mêmes caractéristiques que l'individu parcouru.

De fait, chaque année, de nouveaux individus sont créés avec des caractéristiques issues de la population de base.

6. Les taux de demande d'avantage par mariage

Dans le cadre de la réglementation actuelle :

Selon l'historique fourni, on observe que les demandes se concentrent dans les 3 ans qui suivent les 5 ans de mariages. Nous avons donc considéré qu'une « étrangère » n'effectuant aucune démarche de demande après 8 ans de mariage ne déposerait jamais de dossier.

Ci-dessous les taux de demande appliqués :

Taux demande au cours de la 6 ^{ème} année de mariage	31%
Taux demande au cours de la 7 ^{ème} année de mariage	54%
Taux demande au cours de la 8 ^{ème} année de mariage	30%

Ces taux aboutissent à une demande dans 78% des cas.

Dans le cadre de la réglementation envisagée :

Les même taux ont été appliqués aux hommes pour le scénario 1 :

Taux demande au cours de la 11 ^{ème} année de mariage	31%
Taux demande au cours de la 12 ^{ème} année de mariage	54%
Taux demande au cours de la 13 ^{ème} année de mariage	30%

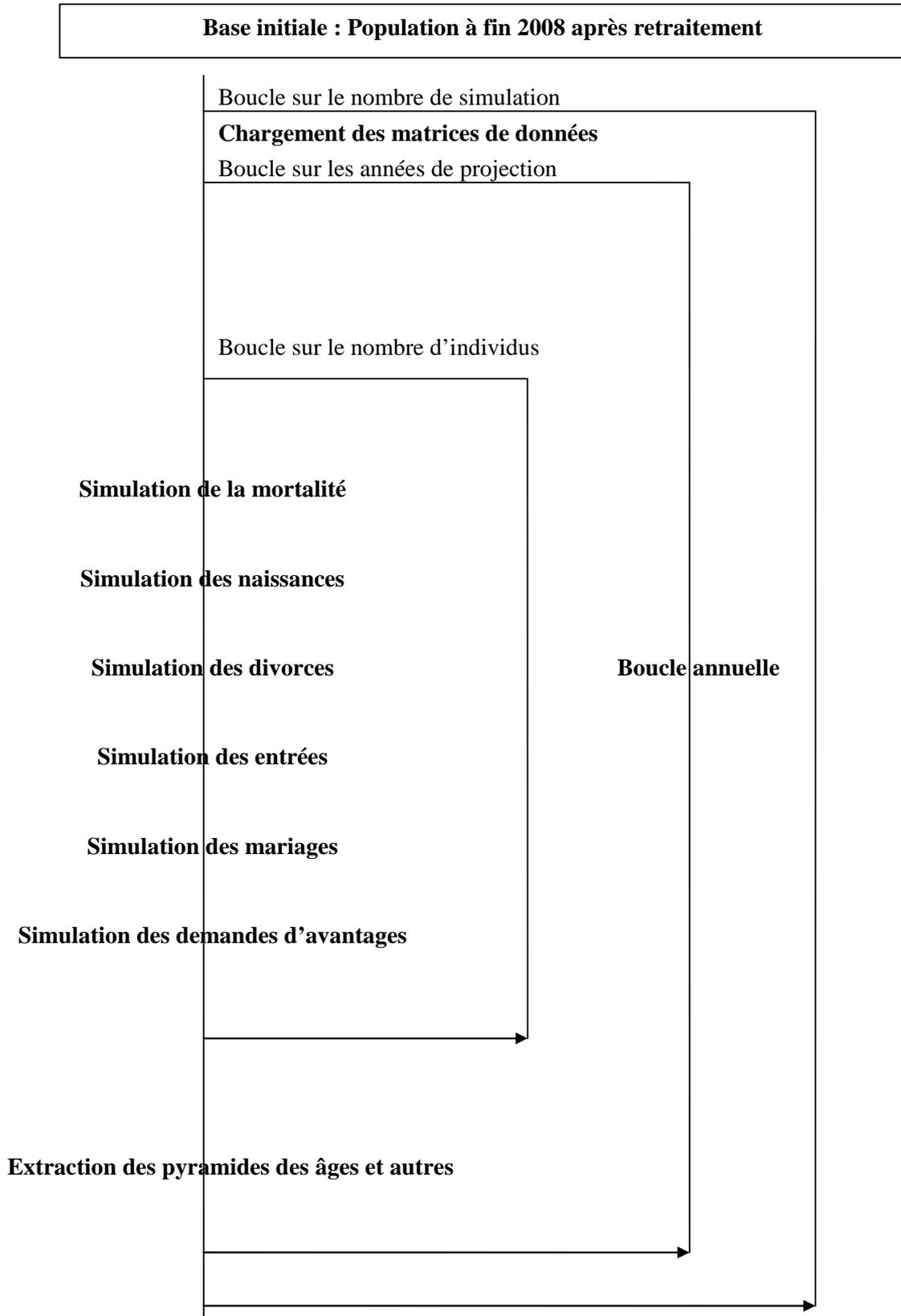
Pour le scénario 2 (hypothèse 100% des « étrangers » déposeront un dossier de demande), il a été considéré les taux suivants :

Taux demande au cours de la 11 ^{ème} année de mariage	40%
Taux demande au cours de la 12 ^{ème} année de mariage	50%
Taux demande au cours de la 13 ^{ème} année de mariage	100%

La partie suivante détaille les premiers résultats de cette étude ainsi que la structure générale du modèle.

C. Programmation du modèle et premiers résultats

1. Présentation du modèle de microsimulation



Nous avons priorisé la mortalité en tant que premier évènement démographique simulé. En effet, il nous semblait plus approprié de faire « décéder » les gens avant de simuler les autres évènements.

La suite du programme simule les naissances, puis les divorces, les entrées, les mariages et enfin les demandes d'avantage. Ce choix est arbitraire et il aurait parfaitement été possible d'utiliser un autre ordre. Nous n'avons pas procédé à d'autres simulations avec une hiérarchisation des évènements différente.

Le modèle utilisé est dynamique et transversal comme définis dans la première partie. Il fonctionne en temps discret. L'état de l'individu est décrit tous les ans (intervalle unitaire) sans aucune information sur les états entre ces deux instants.

Comme nous l'avons expliqué dans la première partie, ces types de modèle ont pour inconvénient un niveau élevé d'exigences en information, un développement informatique important et de longs temps d'exécution ...

Avec cette modélisation, une des faiblesses de la microsimulation est presque immédiatement résolue : l'aléa créé lors de l'échantillonnage. En effet, ici la population restreinte permet de la considérer dans son ensemble (à supposer que les caractéristiques de la population recensée en notre possession soient représentatives des 20% restants).

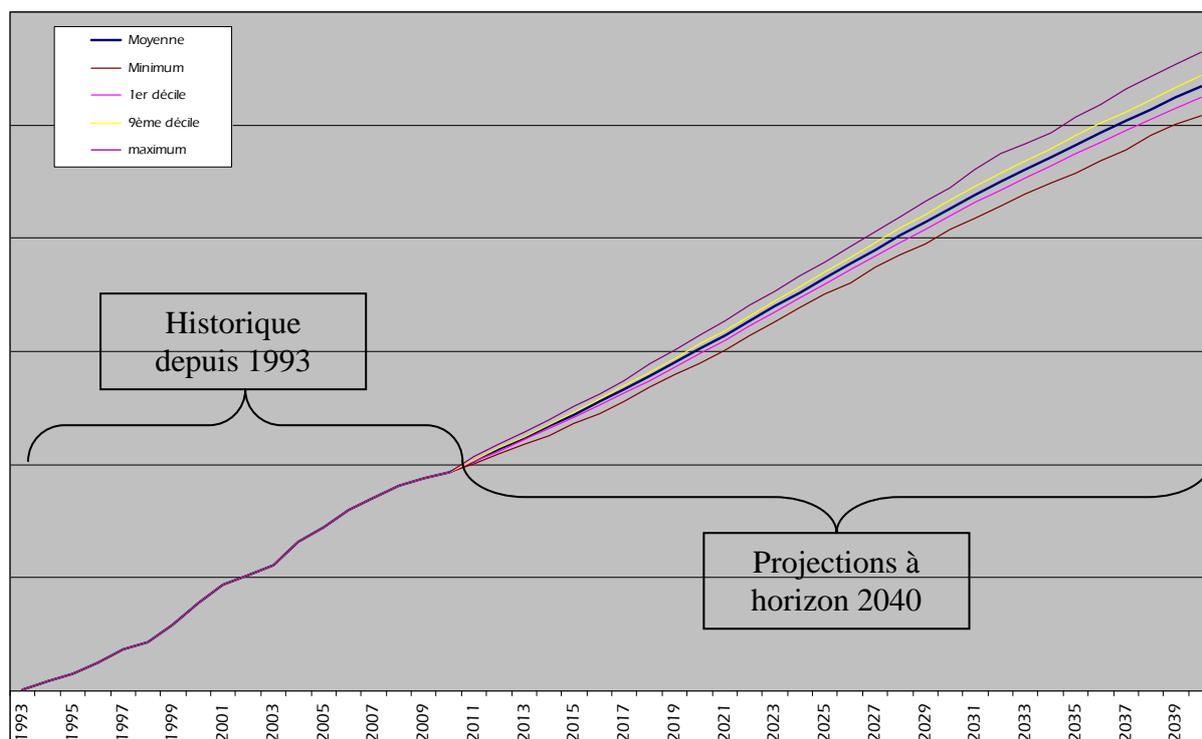
Différentes méthodes sont utilisées pour simuler les différents évènements destinés à faire évoluer la population initiale : probabilité de transition, probabilité d'occurrence des évènements (fécondité, mortalité, mariages), incrémentation automatique de variable (âge, durée de mariage...).

Nos premiers résultats sont décrits dans la sous-partie suivante.

2. Premiers résultats

Nos premières simulations ont d'abord concerné le scénario à réglementation actuelle. Ce dernier se plaçait dans le cadre où le nombre d'entrées simulées est en adéquation avec l'historique fourni.

Ci après le graphique de projection de la population cible à horizon 2040 (10 000 simulations effectuées) :



Ce graphique présente 5 courbes :

- La projection minimum,
- La projection maximum,
- La projection moyenne,
- Le 1^{er} décile des projections,
- Le 9^{ème} décile des projections.

La courbe de projection minimum correspond aux simulations qui donnent pour chaque horizon considéré (2011 à 2040) le nombre minimum d'individus appartenant à la population cible.

La courbe de projection maximum correspond aux simulations qui donnent pour chaque horizon considéré (2011 à 2040) le nombre maximum d'individus appartenant à la population cible.

La courbe de projection moyenne est la moyenne du nombre d'individus simulés entre 2011 et 2040 sur les 10 000 simulations.

La courbe du 1^{er} décile est la courbe qui sépare les 10% de simulations inférieures aux 90% supérieures. Inversement, le 9^{ème} décile est la courbe qui sépare les 10% de simulations supérieures aux 90% inférieures.

On peut remarquer que les 1^{er} et 9^{ème} déciles sont extrêmement proches, même à horizon 2040. En effet, l'écart long terme entre ces deux grandeurs est de l'ordre de moins de 2%.

Les résultats de ces projections sont encourageants : avec nos estimations des lois démographiques on peut remarquer qu'il n'y a pas de « saut » en termes d'effectif entre la partie historique et la partie projection. La pente de croissance est pleinement en accord avec les données en notre possession.

Toutefois, ces projections ne sont pas satisfaisantes en termes de volatilité : les projections sont beaucoup trop lisses et rapprochées les unes par rapport aux autres. Il nous semble en effet que les courbes de minimum, maximum, moyenne, 1^{er} et 9^{ème} décile devraient être moins corrélées et présenter plus de volatilité compte tenu de la méthode employée.

Malgré le caractère « aléatoire » de la microsimulation, nous obtenons une certaine rigidité des résultats qui n'est pas satisfaisante.

Notre première idée a été de mettre en cause le générateur aléatoire. Peut être que ce dernier était « corrompu » ou de mauvaise qualité. Cependant, ce dernier étant considéré comme l'un des plus robustes de nos jours, nous avons écarté cette piste.

Du fait de la pente concordante avec les historiques fournis et la validation de ces lois en amont, nous n'avons pas remis en cause les lois démographiques utilisées.

La méthodologie dans son ensemble est peut être problématique : simuler 6 lois démographiques en parallèle de façon plus ou moins indépendantes a peut être pour conséquence de « figer » les résultats disponibles.

Notre constat s'est donc porté sur les lois démographiques utilisées. En effet, ces dernières, bien que valides, ne tiennent pas compte des variations annuelles observées de l'ordre de 3 à 4% : ces lois sont construites afin de correspondre « en moyenne » à l'évolution de la population et « figent » donc les résultats.

A la lumière de ces constatations nous avons donc introduit dans le modèle des « perturbations » aléatoires afin d'évaser le cône de projection.

Diverses solutions ont été envisagées. L'une d'entre elle consiste à majorer ou minorer directement les lois démographiques au sein du programme. C'est cette méthode qui a été retenue pour cette étude. Cette dernière trouve sa justification dans le fait que chaque année, il est observé des « variations » de la démographie comme expliqué précédemment.

La partie suivante détaille les améliorations du modèle qui ont été apportées ainsi que les caractéristiques du générateur aléatoire utilisé.

Partie III : Amélioration du modèle

Afin d'évaser le cône des projections, nous souhaitons modéliser un taux r qui représente, en pourcentage, la variation annuelle des probabilités de chacune des lois démographiques déjà citées.

Les modèles de taux à un facteur semblent adaptés à l'étude car simples à implémenter et incorporent un retour à la moyenne qui peut être contrôlé.

1. Modèle de taux à un facteur

Les modèles de taux à un facteur sont appréciés car ils incluent dans leur modélisation un retour à la moyenne (qui peut être apprécié).

De manière générale, l'équation stochastique régissant ces taux sous la probabilité historique P se met sous la forme :

$$dr_t = \mu(t, r_t)dt + \sigma(t, r_t)dW_t$$

En appliquant le théorème de Girsanov à $\varphi(t) = -\lambda(t)$ on obtient l'équation suivante :

$$dr_t = (\mu(t, r_t) - \lambda(t)\sigma(t, r_t))dt + \sigma(t, r_t)dW_t^Q$$

Où W_t^Q est un mouvement Brownien sous la probabilité Q .

De cette équation différentielle stochastique découlent de nombreux modèles de taux qui sont largement repris (modèle de Cox Ingeroll Ross, modèle de Hull White...).

Nous nous sommes intéressés à des modèles particuliers. En effet, certains empêchent au taux de prendre des valeurs négatives, voir même de s'annuler (le modèle CIR avec des paramètres soigneusement choisis par exemple).

Nous avons donc considéré le modèle d'Ornstein-Uhlenbeck, modèle de taux à un facteur qui rend possible le fait que le taux r prenne des valeurs négatives.

2. Le modèle d'Ornstein-Uhlenbeck

L'équation différentielle stochastique de ce modèle est la suivante :

$$dr_t = \lambda(\mu - r_t)dt + \sigma dW_t$$

λ est la vitesse de retour à la moyenne qui permet de déterminer le temps moyen de retour à la moyenne long terme,

μ est la moyenne de long terme,

σ la volatilité,

W_t un mouvement brownien standard.

L'introduction de ce modèle dans nos simulations tient du fait que nous voulons évaser nos cônes de projections à l'aide de perturbations sur les lois démographique. Cependant, afin d'être en accord avec l'estimation de ces dernières, nous devons choisir une moyenne long terme nulle pour le taux r . En effet, prendre une valeur non nulle pour la moyenne long terme signifierait que les lois démographiques utilisées ne sont pas « correctes ».

L'équation différentielle devient donc :

$$\begin{cases} dr_t = -\lambda r_t dt + \sigma dW_t \\ r_0 = r \end{cases}$$

Si nous appliquons le lemme d'Itô à $Y_t = e^{\lambda t} r_t$ nous obtenons :

$$\begin{cases} dY_t = \lambda e^{\lambda t} r_t dt + e^{\lambda t} dr_t \\ dY_t = \lambda e^{\lambda t} r_t dt + e^{\lambda t} (-\lambda r_t dt + \sigma dW_t) \\ dY_t = e^{\lambda t} \sigma dW_t \end{cases}$$

Alors, en intégrant entre 0 et t , nous obtenons :

$$\int_0^t dY_u = Y_t - Y_0 = e^{\lambda t} r_t - r_0 = \int_0^t e^{\lambda u} \sigma dW_u$$

$$\text{Soit, } r_t = r_0 + \sigma \int_0^t e^{\lambda(u-t)} dW_u$$

On peut donc déduire que $r_t \approx N(r_0; \sigma^2 \frac{1-e^{-2\lambda t}}{2\lambda})$.

En effet, $\int_0^t e^{-\lambda(u-t)} dW_u$ est une intégrale de Wiener qui est une gaussienne à moyenne nulle.

On a $E(r_t) = r$

$$\text{De plus, } \text{Var}(r_t) = E\left[\left(\sigma \int_0^t e^{\lambda(u-t)} dW_u\right)^2\right] = \sigma^2 \int_0^t e^{2\lambda(u-t)} du = \sigma^2 \frac{1-e^{-2\lambda t}}{2\lambda}$$

On peut donc discrétiser selon le schéma d'Euler :

$$r_{t+\Delta t} = r_t e^{-\lambda \Delta t} + \sigma \sqrt{\frac{1-e^{-2\lambda \Delta t}}{2\lambda}} \varepsilon$$

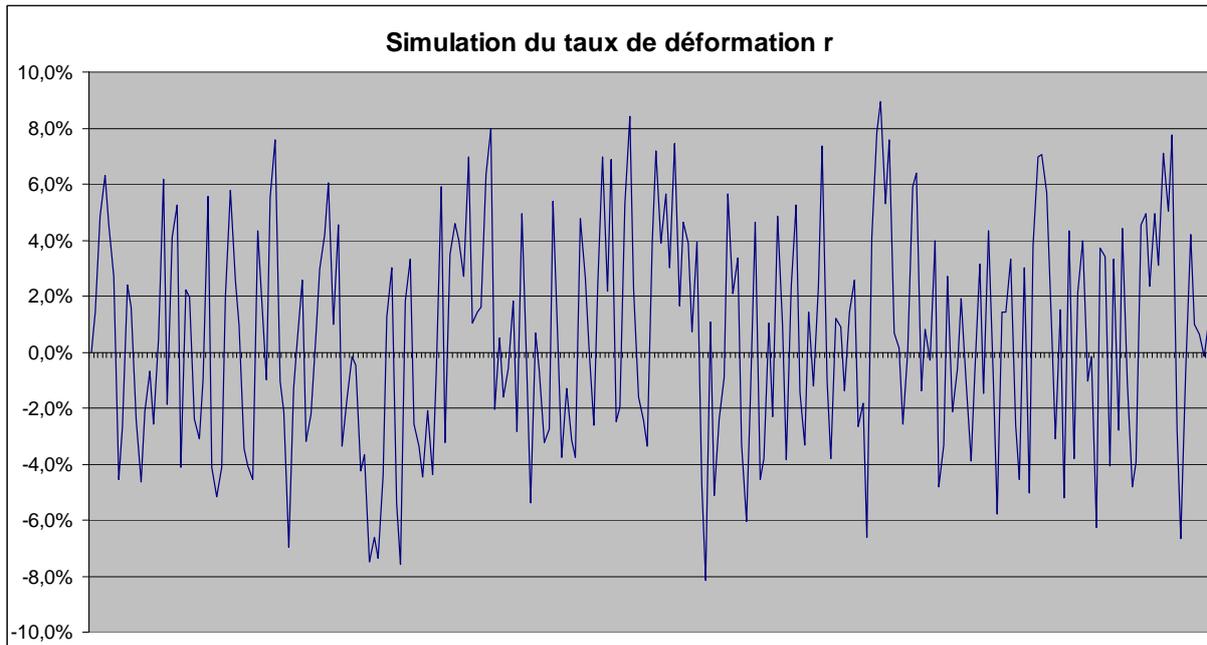
Où ε suit une $N(0,1)$

Nous avons alors le moyen de simuler ce processus AR(1) au moyen d'une valeur initiale r_0 , d'une volatilité σ et d'un coefficient de retour à la moyenne λ .

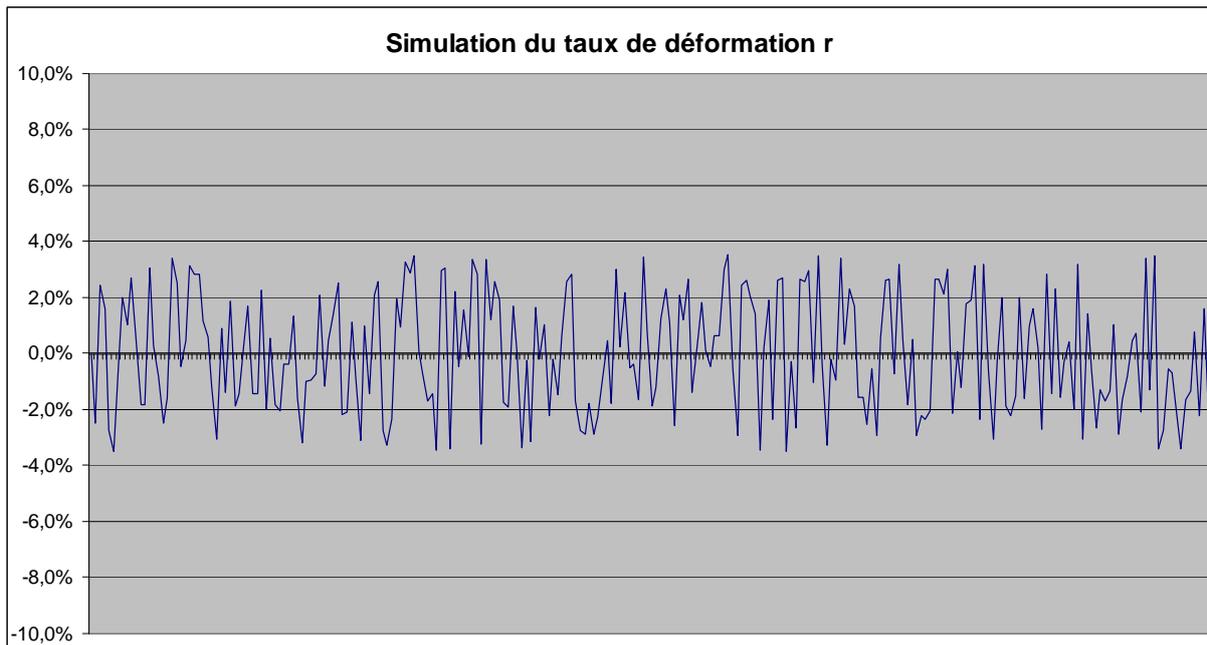
Afin d'être cohérent avec l'analyse des lois démographiques, nous choisissons une valeur nulle pour r_0 et une volatilité $\sigma = 10\%$.

Concernant le coefficient de retour à la moyenne λ , son appréciation est plus délicate. Comme son nom l'indique, plus il sera important et plus la « force de rappel » vers la moyenne sera forte. En d'autres termes, si ce dernier est élevé le taux modélisé aura tendance à moins s'écarter de la moyenne et réciproquement.

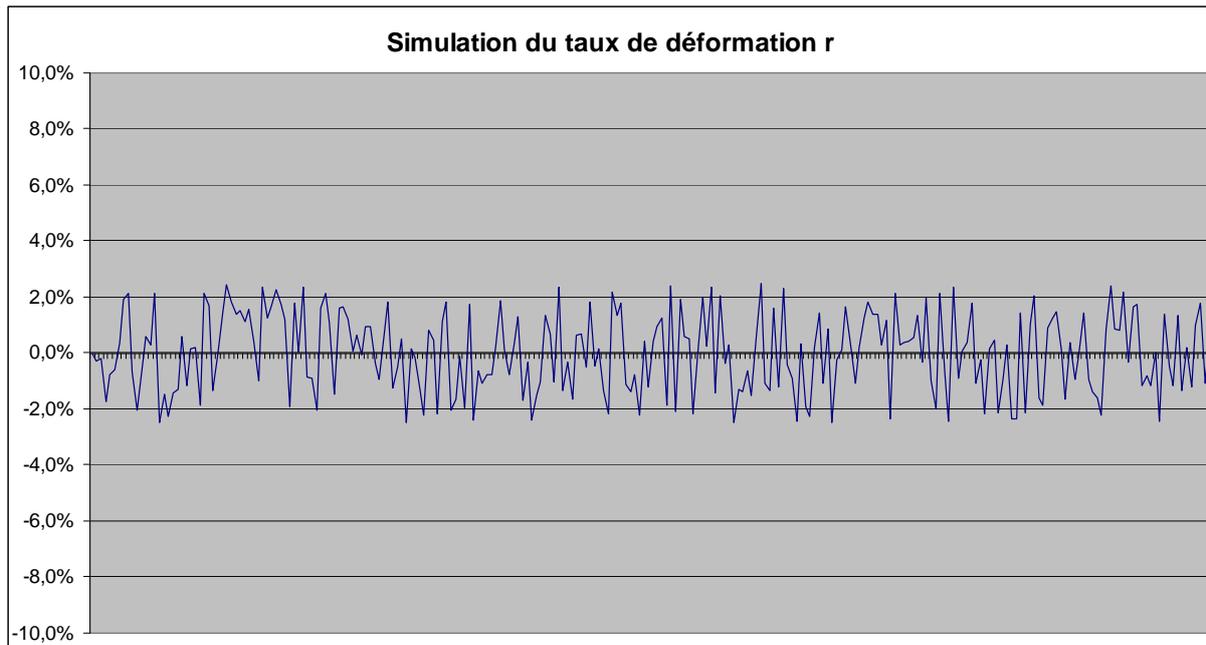
Ci après 3 exemples de trajectoire de modélisations du taux r pour horizon 250 avec un pas de 1 :



$$\sigma = 10\%, r = 0, \lambda = 1$$



$$\sigma = 10\%, r = 0, \lambda = 4$$



$$\sigma = 10\%, r = 0, \lambda = 8$$

On peut effectivement remarquer que plus le coefficient lambda est élevé et moins le taux modélisé prend des valeurs « extrêmes ».

Le premier graphique montre que prendre un coefficient trop faible pour λ peut rendre le taux avoisinant parfois les 9% ce qui est grossièrement inadapté pour cette étude.

Après divers tests effectués au sein du programme de microsimulation, nous avons finalement retenu les paramètres $\sigma = 10\%$, $r = 0$, $\lambda = 4$.

Nous avons donc simulé ce taux de sorte que chaque début d'année au sein de la simulation, la valeur du taux r définie la majoration ou la minoration des taux démographique utilisés.

2. Le générateur aléatoire utilisé

Une attention toute particulière a été portée quant au générateur aléatoire. En effet, les caractéristiques de celui de Visual Basic 2005 sont devenues obsolètes de nos jours.

Celui déjà implémenté utilise la méthode linéaire congruentielle pour générer des nombres aléatoires avec pour construction :

$$X_{n+1} = (X_n * a + c) \text{ MOD } (2^{24})$$

Avec $X_0 = 327680$, $a = 1140671485$ et $c = 12820163$ où l'opérateur MOD retourne le reste de la division euclidienne de $(X_n * a + c)$ par 2^{24} qui est donc égal à X_{n+1} .

En divisant X_i par 2^{24} , on obtient alors un nombre sur] 0;1[. La période maximale est donc de 2^{24} ce qui est grossièrement inadapté de nos jours.

Un autre générateur aléatoire a donc été utilisé dans le but de mieux correspondre au nombre d'appel fait à cette fonction. Il s'agit du générateur Mersenne Twister (développé par Makoto Matsumoto et Takuji Nishimura en 1997. La version utilisée dans cette étude est codée en Visual Basic) qui a passé avec succès différents tests statistiques. La description du générateur est donnée en Annexe 1.

Partie III : Résultats des projections à réglementations distinctes

En préalable à la présentation des résultats obtenus, nous souhaitons rappeler que les projections ont été réalisées en appliquant à la population cible des lois représentant les comportements de cette population sur les dix à quinze années passées.

Nous n'avons pas introduit dans ces projections de possibles changements de comportement, notamment en matière de mariage ou de natalité.

Si ces comportements devaient changer dans le futur, ces projections pourraient s'avérer erronées.

Ce préalable étant posé, nous avons réalisé, à partir de la population initiale à fin 2008, 10 000 projections à horizon 2040. La population au 31/12/2010 étant connue, nous avons recalé ces simulations au 31/12/2010.

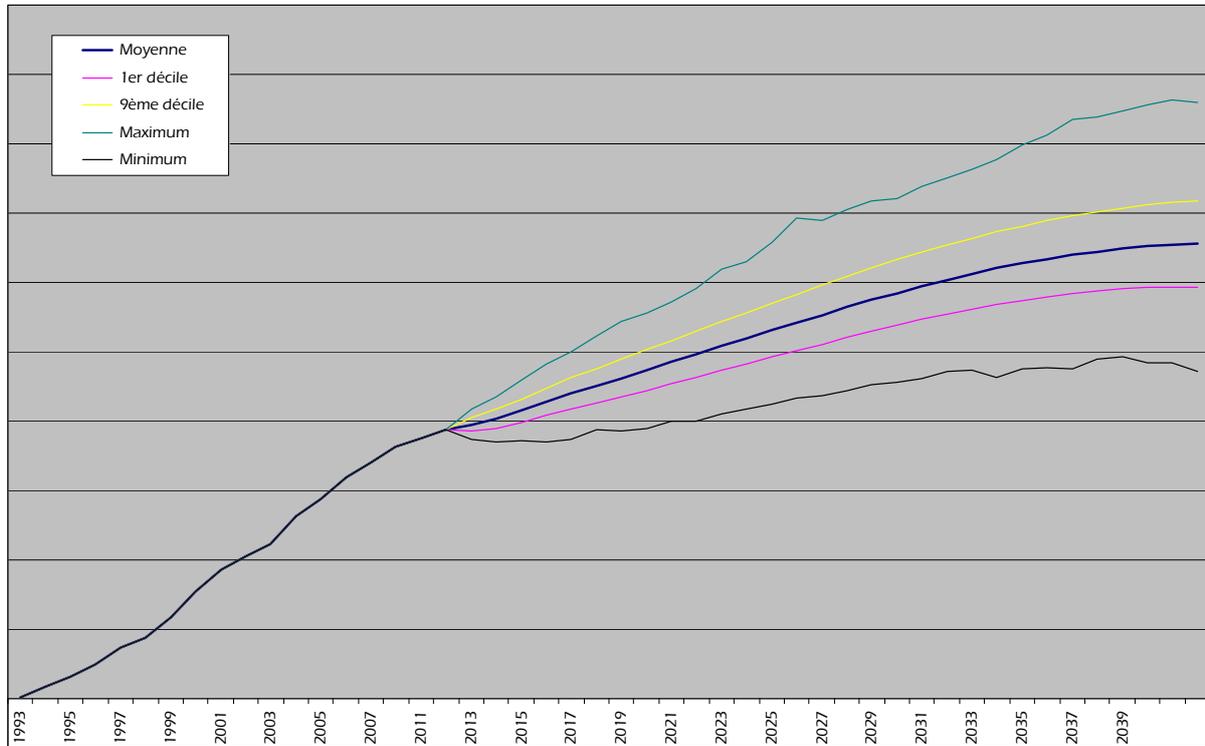
Nous rappelons également que le scénario 1 est le scénario à réglementation envisagée où le comportement des hommes est identique à celui des femmes observé jusqu'à présent en ce qui concerne les demandes d'avantage.

Le scénario 2, dit « catastrophe », part du postulat que tout les hommes « étrangers » mariés à une personne de la population cible pendant au moins 10 ans feront une demande d'avantage dans les trois ans qui suivent.

A. Les résultats des projections à réglementation actuelle

Les résultats sont présentés graphiquement ci-après :

Projection de la population, sous la réglementation actuelle



On remarque que la pente des projections est moins importante que lors des premiers résultats et que celle de l'historique fourni. Ceci tient du fait qu'au lieu de considérer une loi de n entrées annuelles comme le donne l'historique, il n'a été considéré qu'une loi simulant approximativement $n/2$ entrées annuelles comme expliqué auparavant.

Les résultats sont résumés dans le tableau ci-après. Ils sont exprimés en pourcentage de croissance par rapport à la population de base connue en 2010 :

	2020	2030	2040
Projection moyenne	6,58%	13,01%	16,10%
Nombre minimum	0,75%	5,13%	5,06%
Nombre maximum	12,30%	21,88%	28,30%
1er décile	4,58%	10,03%	12,38%
9ème décile	8,57%	15,99%	19,86%

Ainsi, d'après ces résultats, la population pourrait s'accroître, à horizon 2020, de 0.75% à 12.30%, mais plus vraisemblablement de 4.58% à 8.57% (en enlevant les simulations extrêmes), avec une moyenne de croissance probable à 6.58%.

De même, la population pourrait vraisemblablement s'accroître de 10.03% à 15.99% à horizon 2030 et de 12.38% à 19.86% à horizon 2040.

La répartition des simulations réalisées en fonction des effectifs projetés par horizon est présentée graphiquement en page suivante.

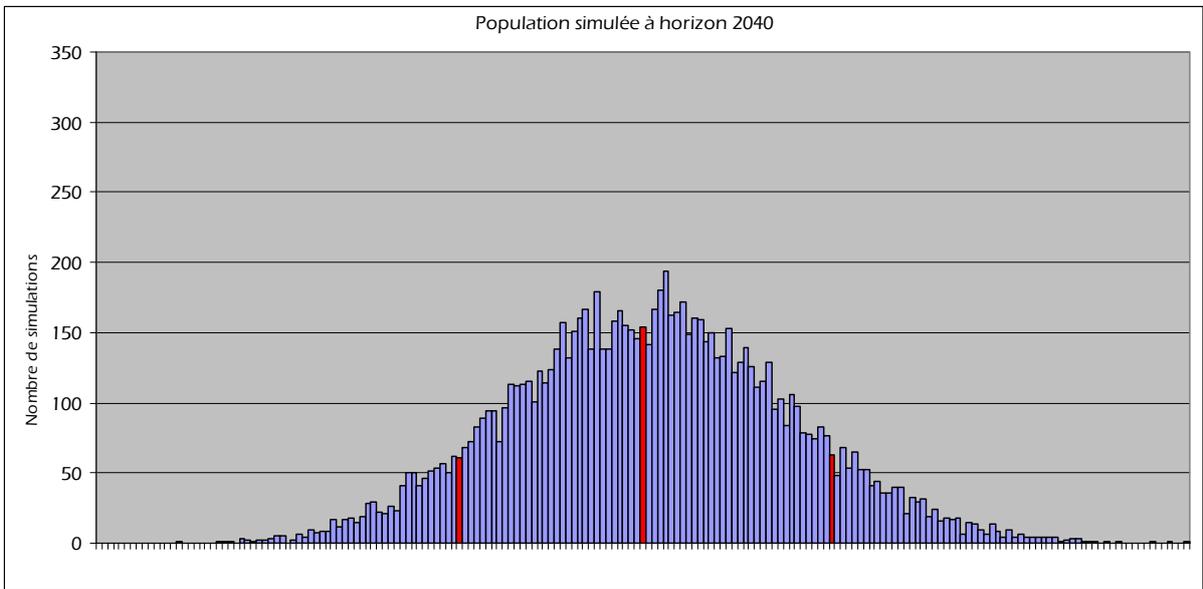
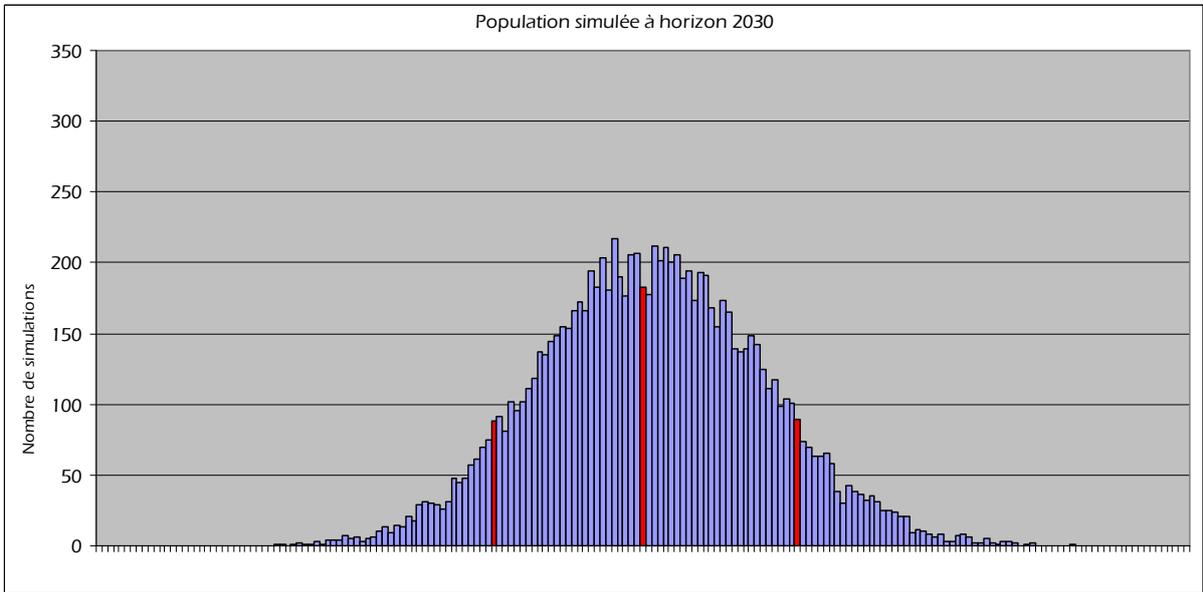
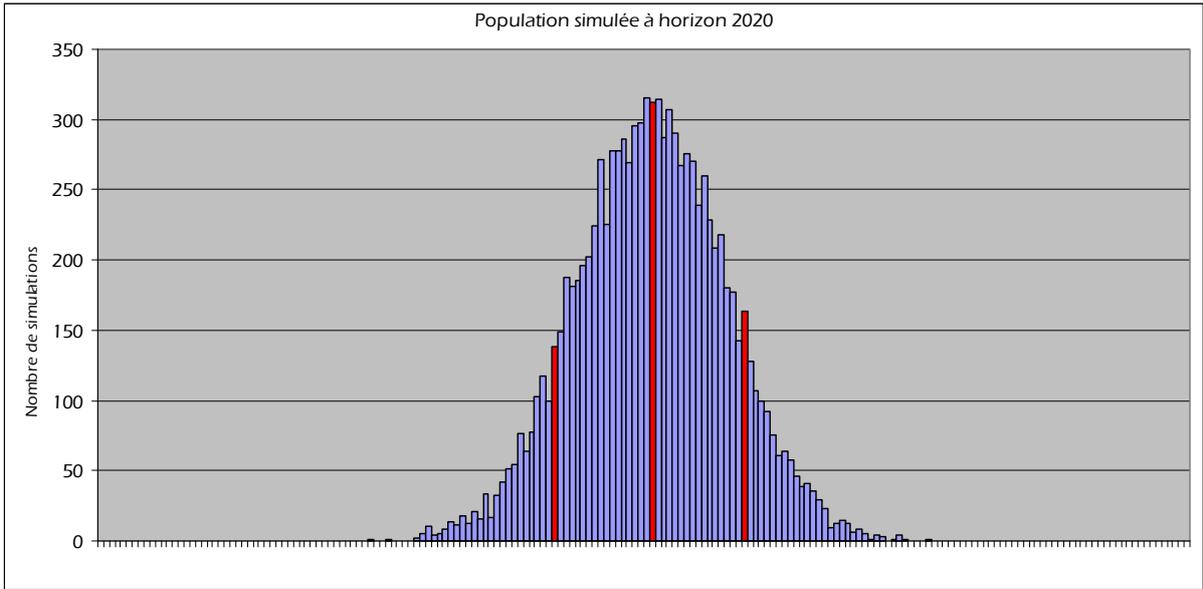
L'échelle et l'amplitude étant identiques pour les trois graphiques (aux horizons 2020, 2030 et 2040), il apparaît clairement que le degré d'incertitude des projections augmente avec la durée de la projection – sur chacun des graphiques, nous avons fait apparaître en rouge la moyenne, et les 1^{er} et 9^{ème} déciles.

Les résultats des projections à l'horizon 2020 sont très concentrés autour de la moyenne.

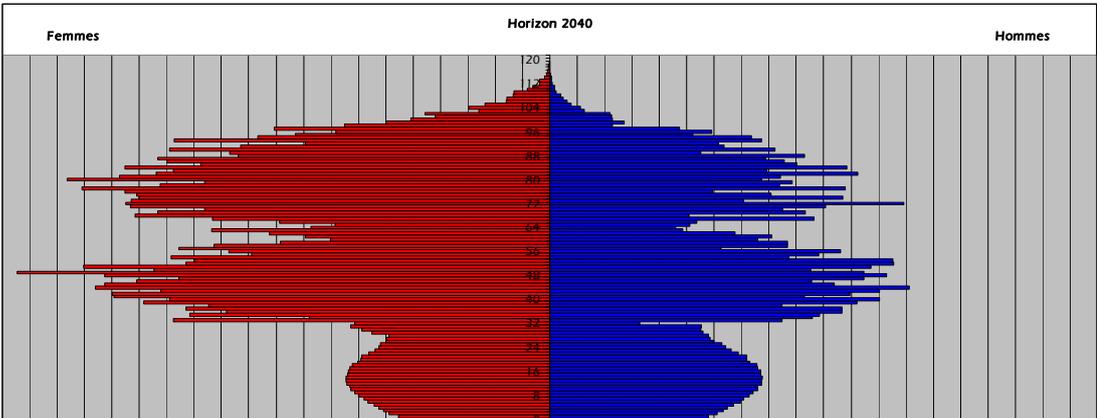
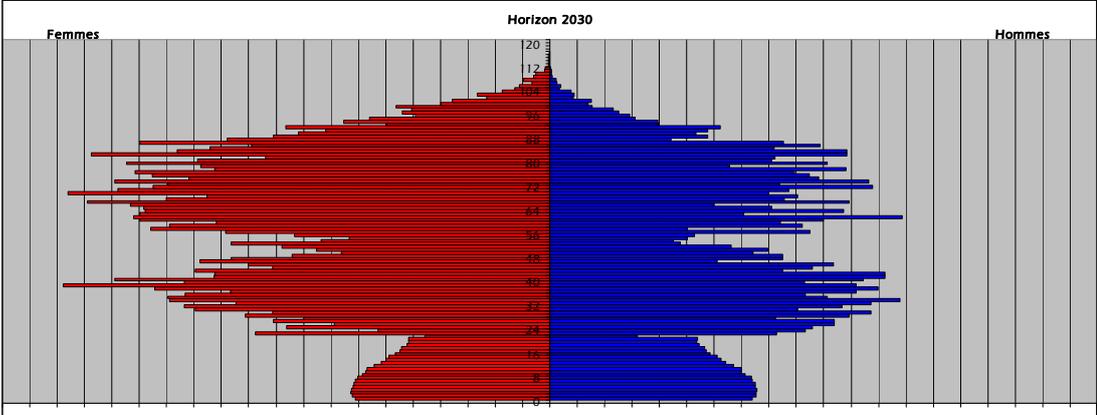
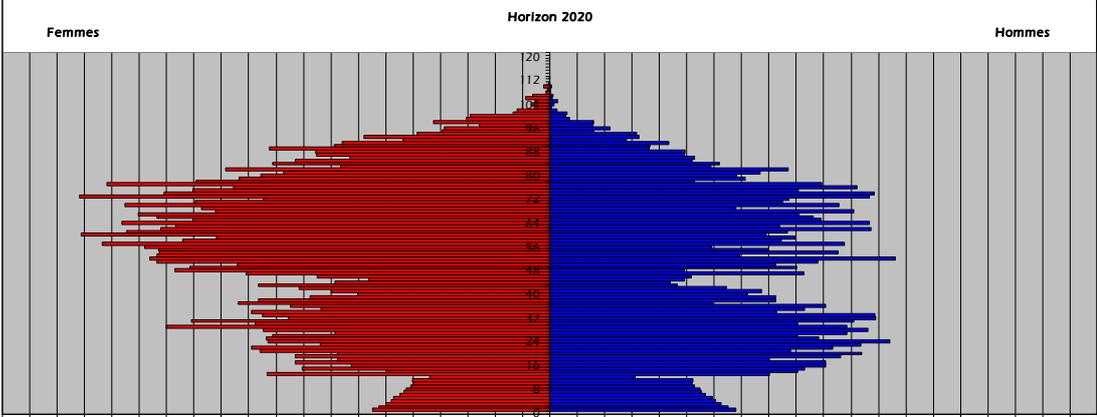
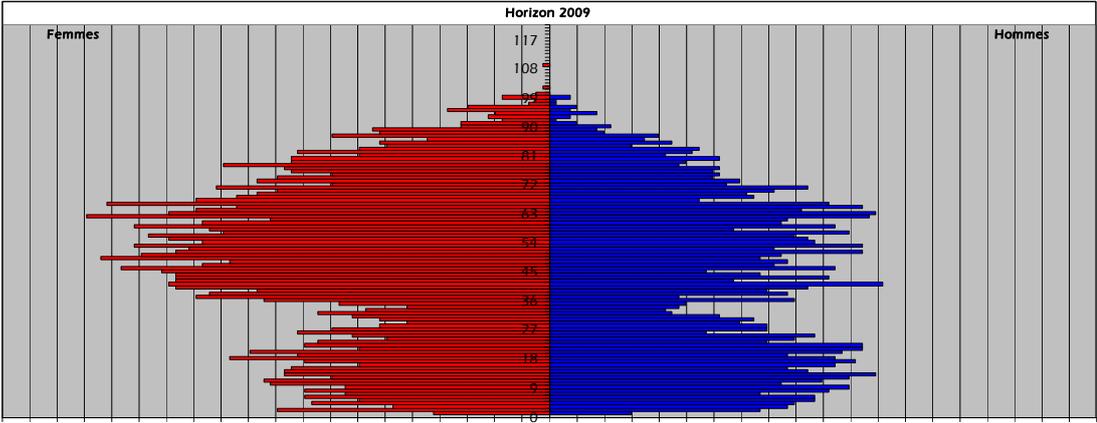
A titre d'information, les écarts types³ mesurés par année de projection du nombre d'individus simulé figurent ci après (en pourcentage par rapport aux projections moyennes):

³ : L'écart type représente l'écart autour de la moyenne qui intègre 2/3 des simulations traitées.

Année de projection	Ecart type mesuré	Année de projection	Ecart type mesuré
2011	0,45%	2026	1,86%
2012	0,63%	2027	1,91%
2013	0,78%	2028	1,97%
2014	0,89%	2029	2,02%
2015	1,01%	2030	2,07%
2016	1,11%	2031	2,11%
2017	1,20%	2032	2,16%
2018	1,30%	2033	2,20%
2019	1,38%	2034	2,24%
2020	1,47%	2035	2,28%
2021	1,54%	2036	2,33%
2022	1,61%	2037	2,37%
2023	1,68%	2038	2,41%
2024	1,75%	2039	2,46%
2025	1,81%	2040	2,51%



Enfin, nous avons extrait les pyramides des âges pour ce scénario :



B. Les résultats des projections à réglementation envisagée

Actuellement, l'avantage n'est accordé qu'aux femmes « étrangères » mariées à un homme de la population cible, après cinq ans de mariage.

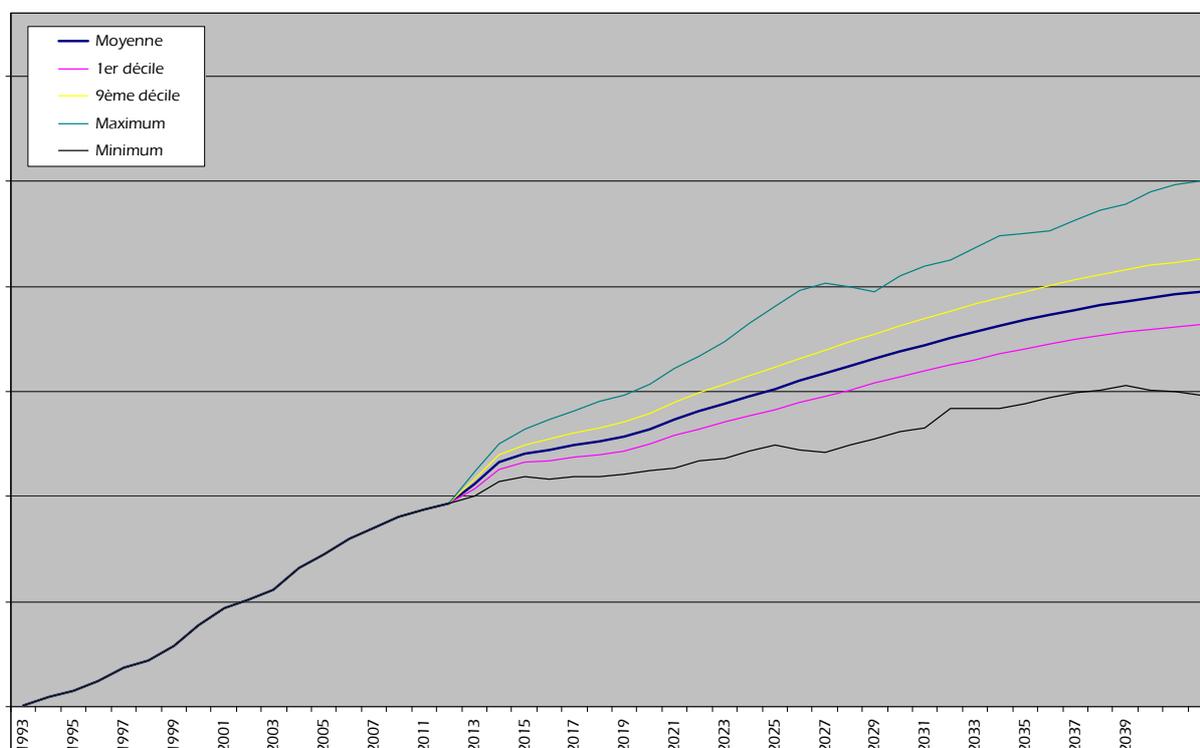
Le changement de réglementation envisagé consiste à ouvrir l'avantage aux hommes mariés à une femme issue de la population cible, mais seulement (tant pour les hommes que pour les femmes) après dix ans de mariage, au lieu de cinq comme actuellement.

1. Scénario 1

Ce scénario prend en compte le changement réglementaire envisagé dès l'horizon 2011 en répliquant, pour les hommes, les comportements observés des femmes quant à leur demande d'avantage sous la règle actuelle.

Les résultats sont présentés graphiquement ci-après :

Projection de la population, sous la réglementation envisagée
Hypothèse de comportement des hommes aligné sur celui des femmes



Les résultats sont résumés dans le tableau ci-après. Ils sont exprimés en pourcentage de croissance par rapport à la population de base connue en 2010 :

	2020	2030	2040
Projection moyenne	10,50%	18,87%	24,16%
Nombre minimum	4,89%	10,79%	12,38%
Nombre maximum	16,81%	27,77%	36,86%
1er décile	8,49%	15,77%	20,41%
9ème décile	12,56%	21,94%	27,97%

Ainsi, d'après ces résultats, la population pourrait s'accroître, à horizon 2020, de 4.89% à 16.81%, mais plus vraisemblablement de 8.49% à 12.46% (en enlevant les simulations extrêmes), avec une moyenne probable de croissance à 10.50%.

De même, la population pourrait vraisemblablement s'accroître de 15.77% à 21.94% à horizon 2030 et de 20.41% à 27.97% à horizon 2040.

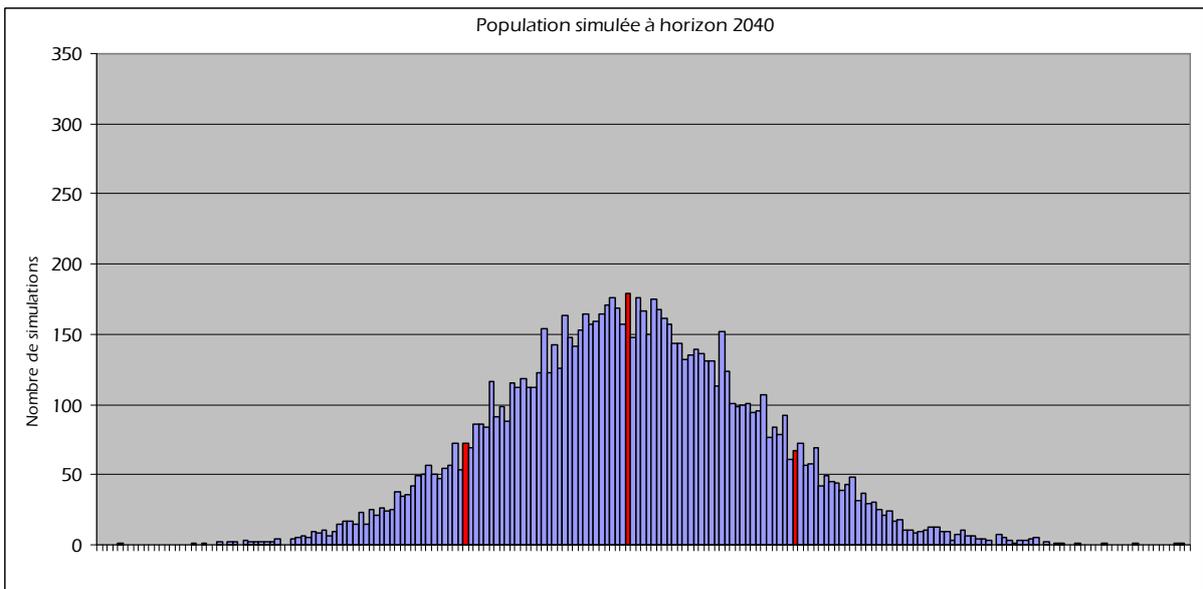
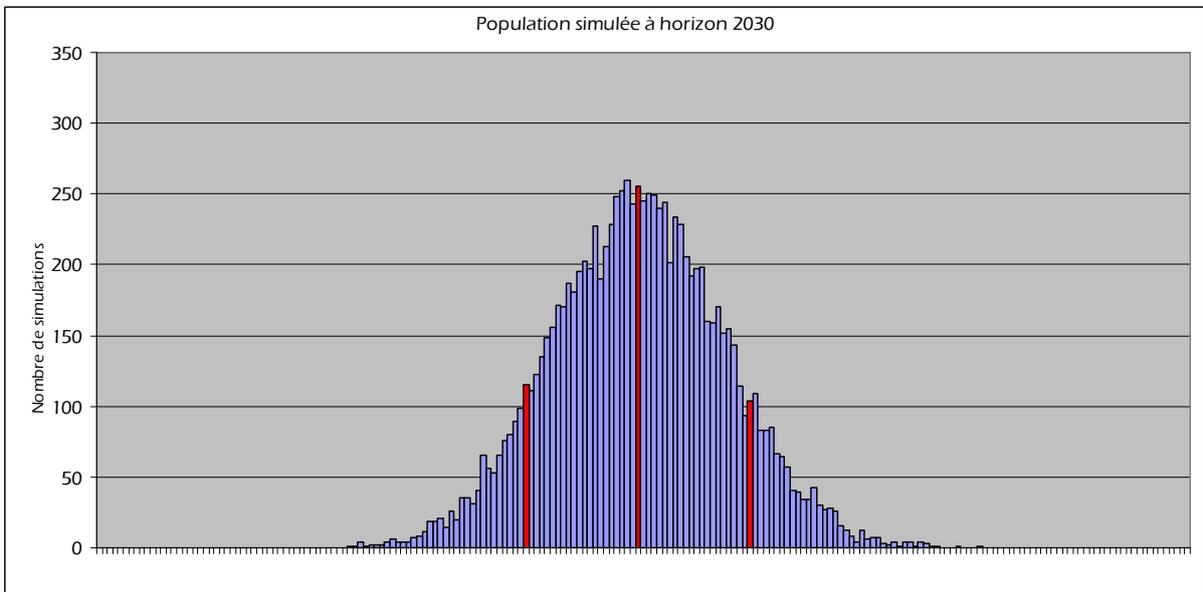
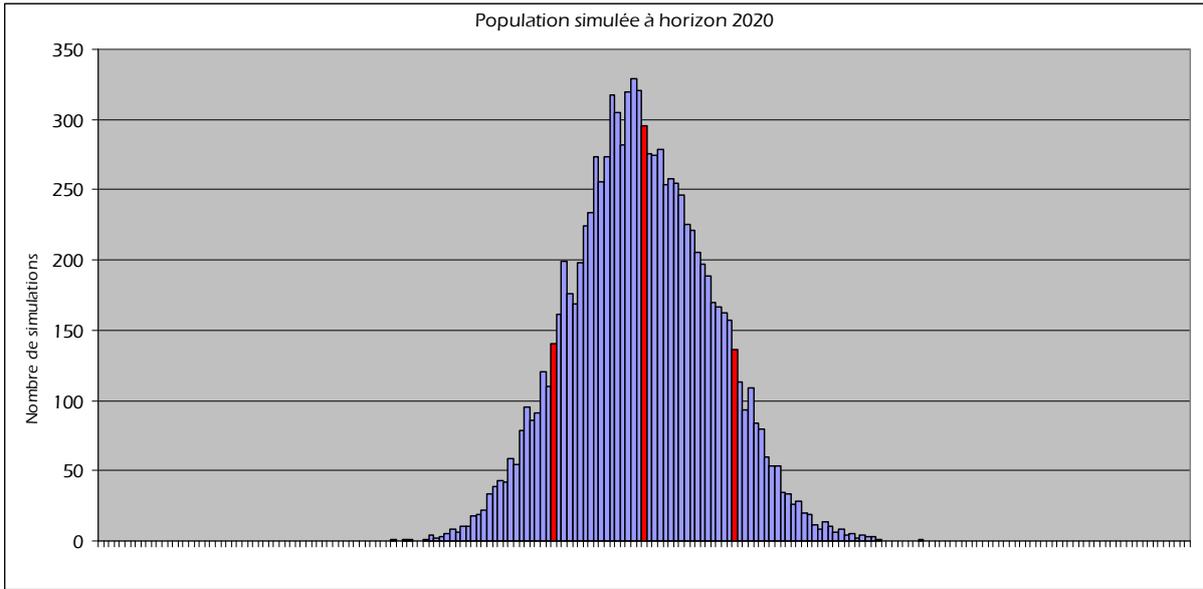
A titre d'information, sont présentés ci-après les écarts types mesurés par année de projection du nombre d'individus simulé (en pourcentage par rapport aux projections moyennes) :

Année de projection	Ecart type mesuré	Année de projection	Ecart type mesuré
2011	0,45%	2026	1,84%
2012	0,62%	2027	1,89%
2013	0,75%	2028	1,94%
2014	0,88%	2029	1,99%
2015	1,00%	2030	2,03%
2016	1,09%	2031	2,07%
2017	1,19%	2032	2,11%
2018	1,28%	2033	2,13%
2019	1,35%	2034	2,16%
2020	1,43%	2035	2,20%
2021	1,52%	2036	2,24%
2022	1,59%	2037	2,27%
2023	1,66%	2038	2,30%
2024	1,72%	2039	2,34%
2025	1,79%	2040	2,38%

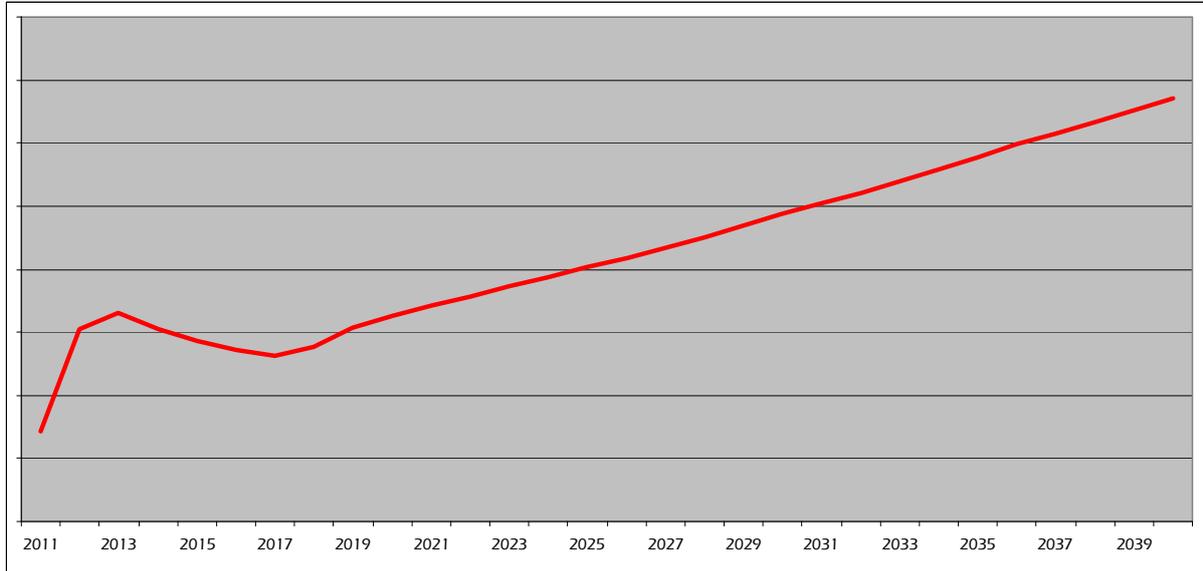
La répartition des simulations réalisées sous ce scénario 1, en fonction des effectifs projetés par horizon est présentée graphiquement en page suivante.

Par rapport à la réglementation actuelle, la population augmenterait en moyenne de :

- 3.68% en 2020,
- 5.18% en 2030,
- 6.94% en 2040.



Augmentation moyenne des effectifs
entre les projections à réglementation actuelle et les projections à réglementation
modifiée
avec hypothèse de comportement des hommes aligné sur celui des femmes



On peut observer une augmentation importante du nombre de personnes ayant perçu l'avantage durant les trois ans⁴ qui suivent l'application de la réglementation envisagée. Il s'agit de « l'effet d'aubaine », le changement législatif ouvrant la possibilité de le demander à tous les hommes ayant déjà dix ans de mariage à la date de promulgation de la loi.

De plus, nous avons considéré que toutes les femmes qui avaient déjà cinq ans de mariage à la date d'adoption de la nouvelle règle pourraient également continuer à demander l'avantage, sans attendre d'avoir dix ans de mariage.

Ainsi, si les hommes ont le même comportement que les femmes, le changement réglementaire pourrait conduire à l'augmentation de la population considérée d'environ 3.50% dans les trois ans de sa promulgation par rapport au scénario à réglementation actuelle.

Après cette première augmentation, on constate une moindre augmentation de la croissance, due au fait que les femmes ayant moins de cinq ans de mariage à la date d'adoption de la

⁴ : Nous avons en effet considéré qu'après l'atteinte de la durée de mariage nécessaire pour demander l'avantage, les demandes étaient effectuées et acceptées sur une période de trois ans.

nouvelle réglementation sont obligées d'attendre cinq ans de plus avant de pouvoir demander l'avantage. Pendant cinq ans donc, aucune femme ne le demandera (sauf celles terminant la procédure entamée, si elles avaient déjà cinq ans de mariage à la date d'adoption de la nouvelle règle).

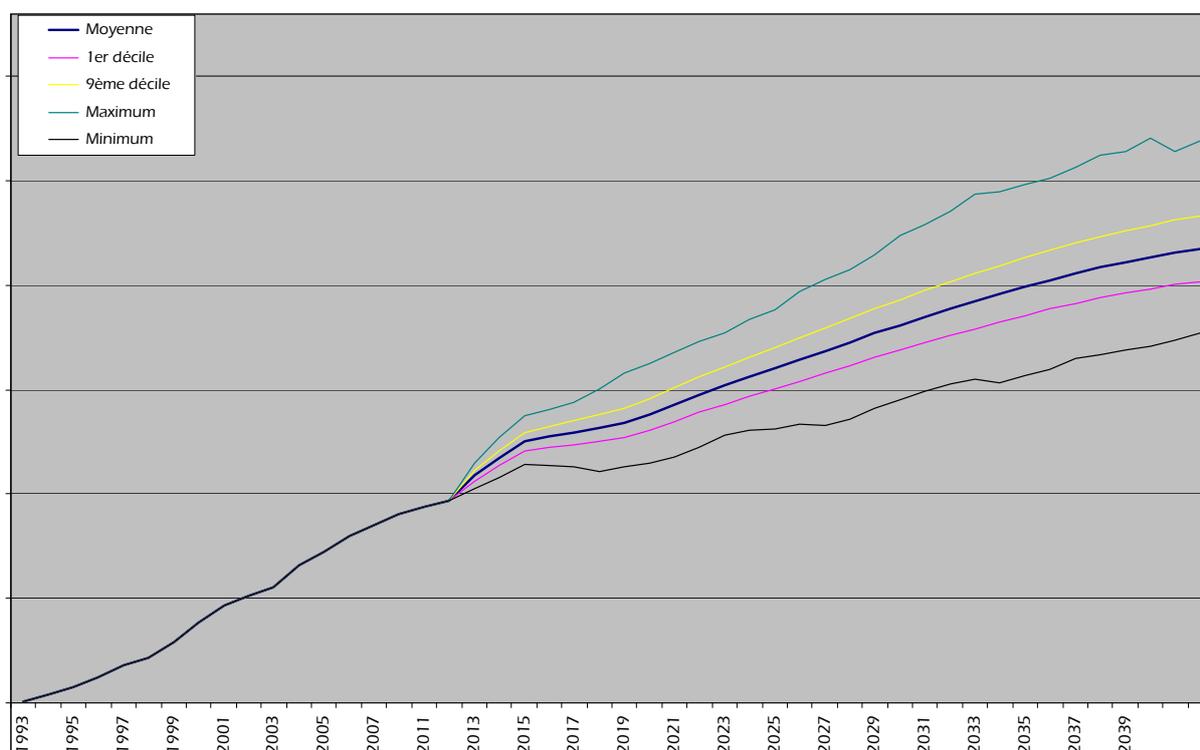
Après cette période transitoire, on revient à un régime « de croisière », où le nombre de nouveaux octroies d'avantage par mariage est un peu plus élevé qu'actuellement. Il n'y a pas deux fois plus de demandes, dans la mesure où, compte tenu des probabilités de divorce retenues dans les simulations, il y a moins de couples qui arrivent à dix ans de mariage qu'à cinq ans.

2. Scénario 2

Ce scénario prend en compte le changement réglementaire envisagé dès l'horizon 2011 avec pour hypothèse que 100% des « étrangers » auront demandé l'avantage passé un délai de 3 ans suivant les 10 ans de mariage requis.

Les résultats sont présentés graphiquement ci-après :

Projection de la population, sous la réglementation envisagée
Hypothèse de 100% de demande d'avantage



Les résultats sont résumés dans le tableau ci-après. Ils sont exprimés en pourcentage de croissance par rapport à la population de base connue en 2010 :

	2020	2030	2040
Projection moyenne	12,22%	22,08%	29,00%
Nombre minimum	6,20%	13,47%	19,32%
Nombre maximum	18,37%	33,33%	41,36%
1er décile	10,17%	18,98%	25,20%
9ème décile	14,25%	25,19%	32,81%

Ainsi, d'après ces résultats, la population pourrait vraisemblablement s'accroître, à horizon 2020, de 10.17% à 14.25% (en enlevant les simulations extrêmes), avec une moyenne probable à 12.22%.

De même, la population pourrait augmenter de 18.98% à 25.19% à horizon 2030 et de 25.20% à 32.81% à horizon 2040.

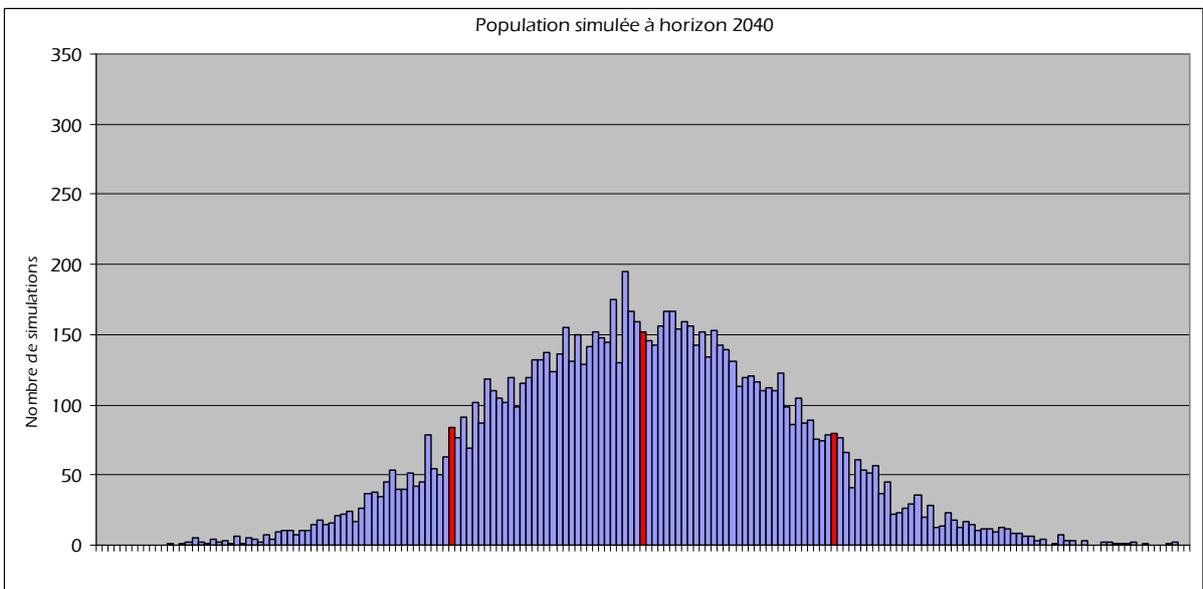
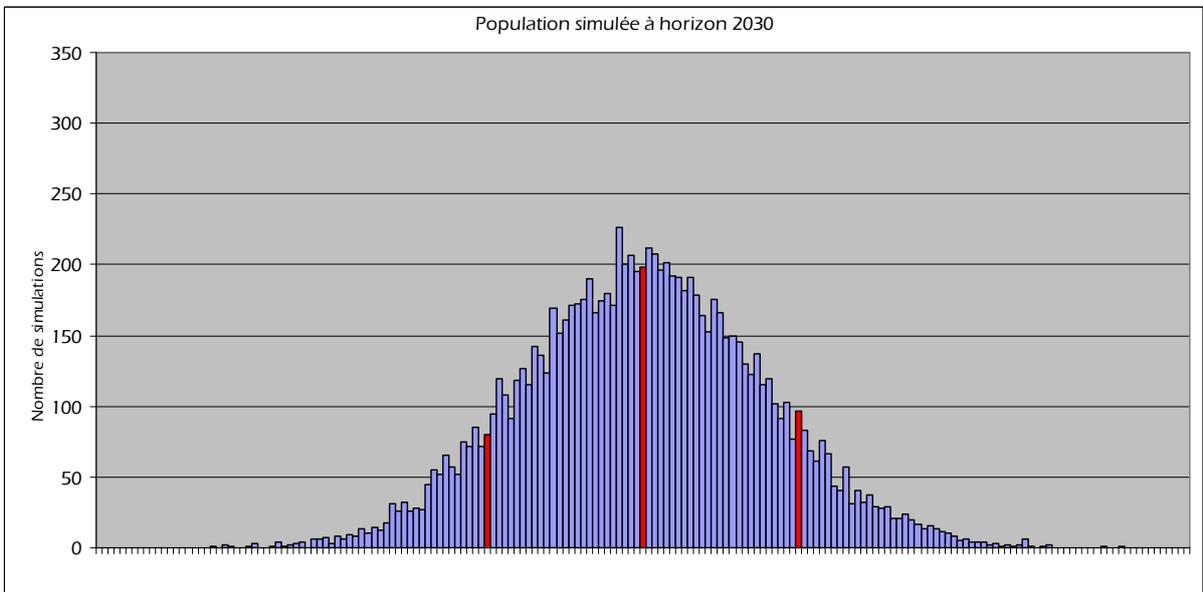
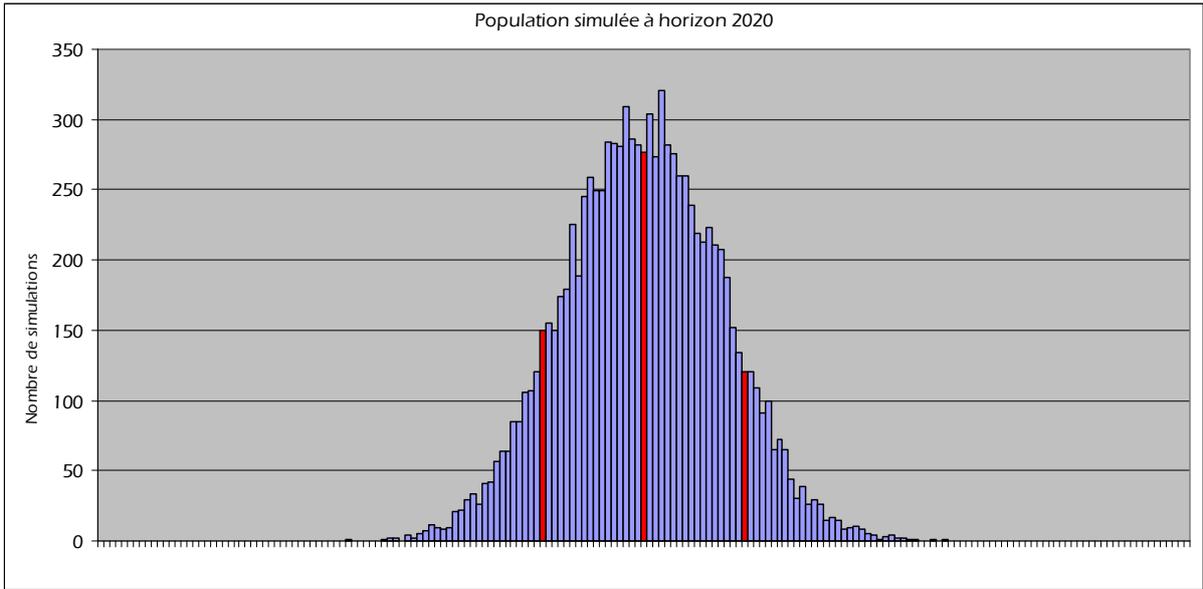
A titre d'information, ci après les écarts types mesuré par année de projection du nombre d'individu simulé :

Année de projection	Ecart type mesuré	Année de projection	Ecart type mesuré
2011	0,45%	2026	2,05%
2012	0,63%	2027	2,09%
2013	0,76%	2028	2,12%
2014	0,89%	2029	2,18%
2015	0,99%	2030	2,24%
2016	1,09%	2031	2,29%
2017	1,19%	2032	2,34%
2018	1,27%	2033	2,37%
2019	1,35%	2034	2,38%
2020	1,42%	2035	2,41%
2021	1,49%	2036	2,43%
2022	1,55%	2037	2,45%
2023	1,61%	2038	2,49%
2024	1,68%	2039	2,51%
2025	1,73%	2040	2,54%

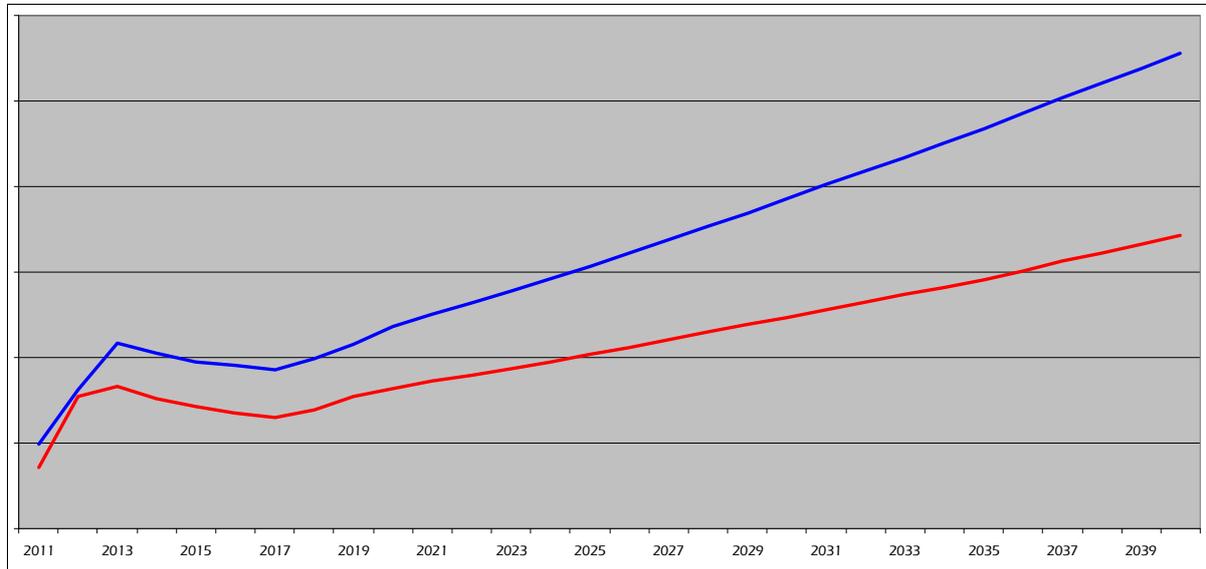
La répartition des simulations réalisées sous ce scénario 2, en fonction des effectifs projetés par horizon est présentée graphiquement en page suivante.

Par rapport à la réglementation actuelle, la population augmenterait en moyenne de :

- 5.29% en 2020,
- 8.03% en 2030,
- 11.11% en 2040.



**Augmentation moyenne des effectifs
entre les projections à réglementation actuelle et les projections à réglementation
modifiée**



On peut observer (courbe bleue) un saut un peu plus important du nombre d'individus durant les trois ans⁵ qui suivent l'application de la réglementation envisagée par rapport au scénario 1 (représenté par la courbe rouge), puis, après la même décroissance due aux femmes ayant moins de cinq ans de mariage à la date de promulgation de la loi, une reprise de croissance également un peu supérieure à celle observée dans le scénario précédent.

Ainsi, si tous les « étrangers » mariés à un individu issu de la population de base depuis plus de 10 ans font une demande d'avantage, le changement réglementaire pourrait conduire à l'augmentation de la population d'environ 4.70% dans les trois ans de sa promulgation (au lieu de 3.50% dans le scénario 1),

D'autres scénarios ont été considérés. Les résultats figurent tous en annexe 3.

⁵ : Nous avons aussi ici considéré qu'après l'atteinte de la durée de mariage nécessaire pour demander l'avantage, les demandes étaient effectuées et acceptées sur une période de trois ans.

Conclusion

Les projections de la population présentées dans ce document ont été réalisées à partir de la population au cours de l'année 2008. Cette population représentait 80 % de la population totale au 31/12/2008.

Les hypothèses utilisées ont été établies à partir de l'observation des comportements de la population au cours des dix dernières années. Nous n'avons pas intégré de déformation de ces hypothèses dans le futur (hormis pour la mortalité, pour laquelle nous avons travaillé avec des tables prospectives, intégrant une amélioration de l'espérance de vie).

Sous ces deux hypothèses fortes, les projections réalisées indiquent une croissance continue de la population, qui, sous la réglementation actuelle, pourrait s'accroître de 13% à horizon 2030 (environ 7% à l'horizon 2020).

L'impact du changement de règle envisagé dépendra du comportement des personnes en matière de demande d'avantage. Si les hommes ont le même comportement que le comportement actuel des femmes, le changement de règle pourrait entraîner une croissance immédiate (dans les trois ans) de la population d'environ 3.50% (effet d'aubaine).

Cet accroissement pourrait aller jusqu'à 4.70%, si tous les « étrangers » (hommes et femmes) mariés à une personne issue de la population de base demandaient l'avantage.

La microsimulation permet de faire certaines choses impossibles en macro-simulation et doit, en définitive, être considéré comme un outil potentiel puissant et important pour réaliser des projections, aussi bien démographiques que non démographiques. Cependant il faut prêter attention aux effets de l'aléa, à la simulation des événements « rares », aux choix des hypothèses et à l'interprétation des résultats qui en découle.

Annexe 1

Le Mersenne Twister a été développé par Makoto Matsumoto et Takuji Nishimura en 1997. La version utilisée est la MT 19937 codée en Visual Basic 2005 qui utilise le nombre premier de Mersenne $2^{19937} - 1$. Il possède les caractéristiques suivantes :

- Il a une période de $2^{19937} - 1$
- Il est 623-équidistribué jusqu'à la précision 32 bits.
- Il est extrêmement rapide
- Il passe les tests DIEHARD.

La version codée en Visual Basic a été développée par Ron Charlton en décembre 2008. L'ensemble du code utilisé est disponible dans le package à [cette](#) adresse de même que les résultats des tests de la batterie DIEHARD (résultats disponibles à [cette](#) adresse). Ces tests sont entre autres, le test de collision et le test d'espacement des anniversaires.

A titre d'informations, voici quelques fonctions codées pour ce générateur :

Fonction	Retourne une valeur dans l'intervalle:
genrand_int32()	[0, 4294967295] (0 à $2^{32}-1$)
genrand_int31()	[0, 2147483647] (0 à $2^{31}-1$)
genrand_real1()	[0.0, 1.0] (0.0 et 1.0 inclus)
genrand_real2()	[0.0, 1.0) = [0.0, 0.9999999997672...] (0.0 included, 1.0 excluded)
genrand_real3()	(0.0, 1.0) = [0.0000000001164..., 0.9999999998836...] (0.0 et 1.0 exclus)
genrand_res53()	[0.0, ~1.0] = [0.0, 1.000000000721774...] (0.0 inclus, ~1.0 inclus)

Ron Charlton a aussi ajouté quelques autres fonctions :

Fonction	Retourne une valeur dans l'intervalle:
Genrand_int32SignedInt()	[-2147483648, 2147483647] (-2 ³¹ à 2 ³¹ -1)
Genrand_real2b()	[0.0, 1.0]=[0, 1-(2*gap)] =[0.0, 0.999999999999990] (*) (0.0 inclus, 1.0 exclus)
Genrand_real2c()	(0.0, 1.0)=[0+(2*gap),1.0]=[1.0e-12, 1.0] (*) (0.0 exclus, 1.0 inclus)
Genrand_real3b()	(0.0, 1.0)=[0+gap, 1-gap] =[5.0e-13, 0.99999999999995] (*) (0.0 et 1.0 exclus)
Genrand_real4b()	[-1.0,1.0]=[-1.0, 1.0] (-1.0 inclus, 1.0 inclus)
Genrand_real5b()	(-1.0,1.0)=[-1.0+(2*gap), 1.0-(2*gap)]= [-0.99999999999990,0.99999999999990] (*) (-1.0 exclus, 1.0 exclus)
Genrand_intMax(bornesup)	[0,bornesup] pour bornesup < 2 ³² (0 à 4294967295 mais <= bornesup)
Genrand_intRange(borneinf,bornesup)	[borneinf,bornesup] pour 0 <= borneinf <= bornesup <= 2 ³² -1 (0 <= borneinf<= bornesup <= 4294967295)
Genrand_int64()	[0,18446744073709551615] (0 à 2 ⁶⁴ -1)

Sur un ordinateur Dell Precision 390 possédant un processeur Intel Core 2 duo à 2.66Gz et 3,5Go de RAM, en compilant avec Visual Basic.NET pour générer 100 000 000 nombres pour chacune des fonctions suivantes, les caractéristiques sont :

Fonction	Temps de bouclage en secondes	Nanosecondes par appel	Appels par seconde
genrand_int32()	1,656	16,56	60 390 000
genrand_real1()	2,031	20,31	49 240 000
genrand_res53()	4,25	42,5	23 530 000
genrand_intMax(1000)	2,343	23,43	42 680 000
genrand_intMax(&H5FFFFFFF)	3,359	33,59	29 770 000
genrand_intRange(10,20)	3,813	38,13	26 230 000
genrand_int64()	3,516	35,16	28 440 000

On peut donc conclure que l'appel à ces différentes fonctions ne pénalise aucunement le temps de bouclage d'un programme, même avec une fréquence élevée.

Annexe 2

A. Natalité

Taux de fécondité selon l'âge de la mère		en %
Âge (1)	2010	
15		0,1
16		0,2
17		0,4
18		0,9
19		1,8
20		2,8
21		3,9
22		5,0
23		6,4
24		8,0
25		9,7
26		11,4
27		12,9
28		14,3
29		14,9
30		15,0
31		14,7
32		13,7
33		12,3
34		10,8
35		9,3
36		7,9
37		6,3
38		5,0
39		3,9
40		3,0
41		2,1
42		1,3
43		0,8
44		0,5
45		0,2
46		0,1
47		0,1
48		0,0
49		0,0
50		0,0

(1) : âge atteint dans l'année.

Champ : France métropolitaine.

Source : Insee, estimations de population et statistiques de l'état civil.

B. Nuptialité

TAUX DE NUPTIALITÉ PAR SEXE, GROUPE D'ÂGES ET ÉTAT MATRIMONIAL ANTÉRIEUR AU MARIAGE

Nombre d'hommes nouveaux mariés pour 100 hommes du groupe d'âges considéré

H - Hommes

Ensemble	2,31
15 à 49 ans	2,35
18 à 59 ans	2,57
18 ou 19 ans	0,03
20 à 24 ans	0,78
25 à 29 ans	4,20
30 à 39 ans	4,36
40 à 49 ans	2,41
50 à 59 ans	1,84
60 ans ou plus	0,691

Nombre de femmes nouvelles mariées pour 100 femmes du groupe d'âges considéré

F- Femmes

Ensemble	1,77
15 à 49	2,64
18 à 59 ans	2,72
15 à 19 ans	0,12
20 à 24 ans	1,87
25 à 29 ans	5,58
30 à 39 ans	4,05
40 à 49 ans	2,04
50 à 59 ans	1,17
60 ans ans ou plus	0,132

Champ : France métropolitaine, territoire au 31 décembre 2008

Source : Insee, état civil

C. Divortialité

Taux de divorce selon la durée du mariage et l'année du divorce

Durée du mariage	pour 1 000 mariages	
	2008	
Moins d'un an		0,6
1 an		6,3
2 ans		14,4
3 ans		21,7
4 ans		25,8
5 ans		27,3
6 ans		25,0
7 ans		22,6
8 ans		21,2
9 ans		19,1
10 ans		19,0
11 ans		17,3
12 ans		16,0
13 ans		15,6
14 ans		14,6
15 ans		13,9
16 ans		13,5
17 ans		12,9
18 ans		11,7
19 ans		11,8
20 ans		10,5
21 ans		10,1
22 ans		9,4
23 ans		9,2
24 ans		8,7
25 ans		7,9
26 ans		7,2
27 ans		6,7
28 ans		5,8
29 ans		5,2

Champ : France métropolitaine.

Sources : ministère de la Justice ; Insee.

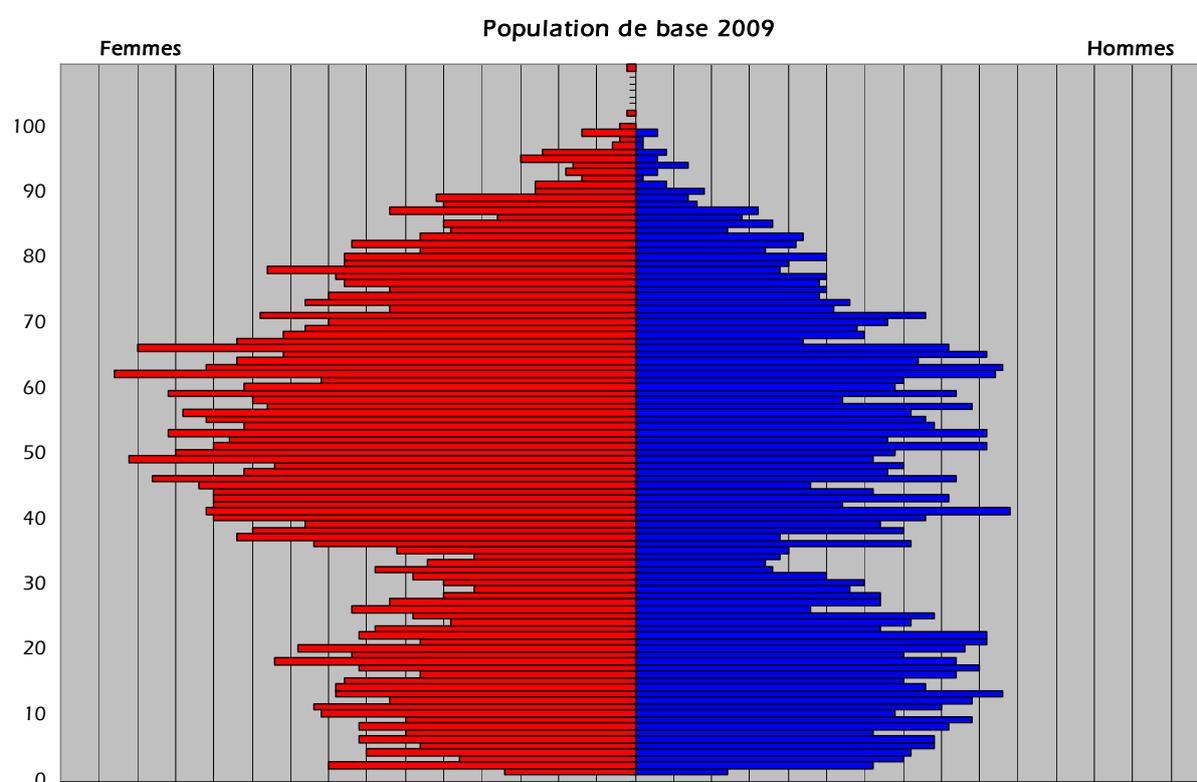
Annexe 3 : Résultats complémentaires

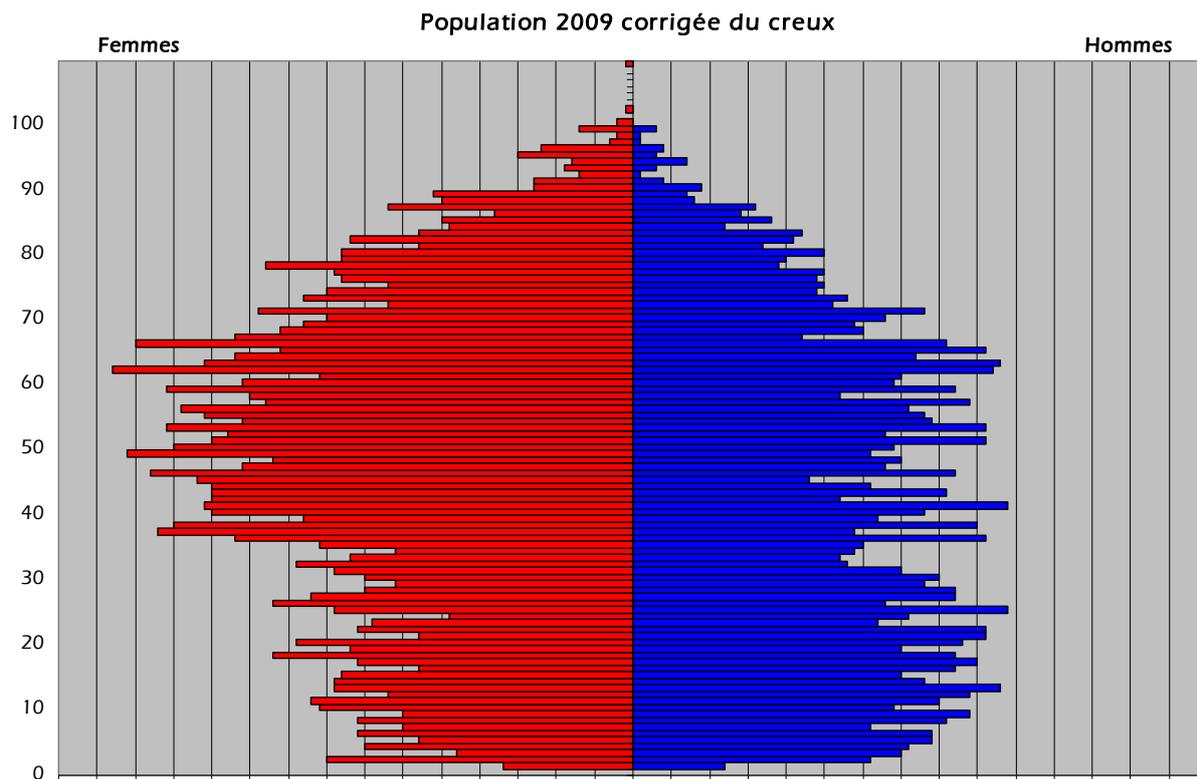
A. Projections après correction du creux de la population autour de 30 ans

Dans le scénario central, nous avons supposé que la population recensée en 2008 était représentative de la population totale, et avons donc corrigé forfaitairement nos projections du rapport de la population totale sur la population recensée.

Cependant, la pyramide des âges des personnes recensées présente une anomalie autour de 30 ans. En conséquence, nous avons réalisé un test, en procédant à une correction des données reçues pour diminuer ce creux démographique. Pour ce faire, nous avons dupliqué des individus autour de 30 ans, répartis équitablement entre hommes et femmes.

Les pyramides des âges avant et après correction sont présentées graphiquement ci-après :





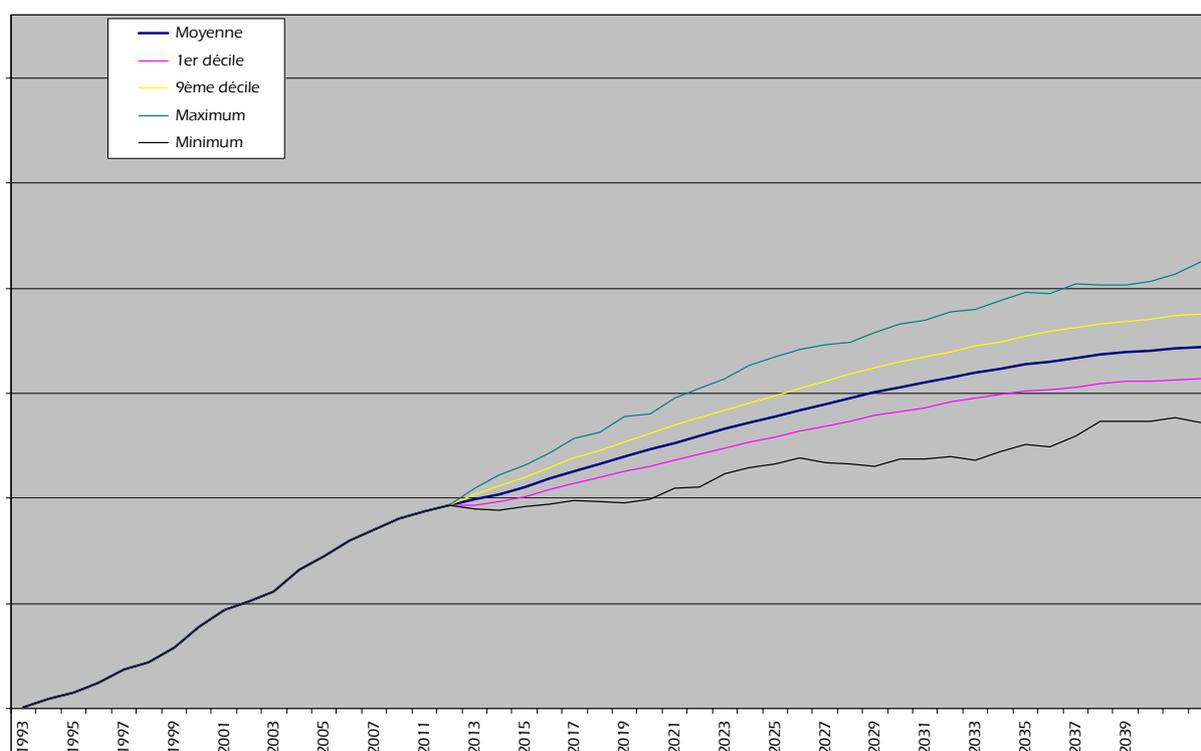
Un léger creux autour de 30 ans a été conservé, nettement moins marqué qu'originellement, mais plus en accord avec les données dont nous disposons.

Les résultats des différents scénarios réalisés avec cette nouvelle population sont présentés ci après.

1. Projections à réglementation actuelle (n/2 entrées)

Les résultats sont présentés graphiquement ci-après :

Projection de la population, sous la réglementation actuelle Avec correction de la population de base



Les résultats sont résumés dans le tableau ci-après. Ils sont exprimés en pourcentage de croissance par rapport à la population de base connue en 2010 :

	2020	2030	2040
Projection moyenne	7,90%	14,57%	17,99%
Nombre minimum	2,04%	5,56%	9,40%
Nombre maximum	13,35%	22,07%	27,84%
1er décile	5,84%	11,78%	14,39%
9ème décile	9,99%	17,55%	21,83%

Ainsi, d'après ces résultats, la population pourrait vraisemblablement s'accroître, à horizon 2020, de 5.84% à 9.99% (en enlevant les simulations extrêmes), avec une moyenne de croissance probable égal à 7.90%.

De même, la population pourrait augmenter de 11.78% à 17.55% à horizon 2030 et de 14.39% à 21.83% à horizon 2040.

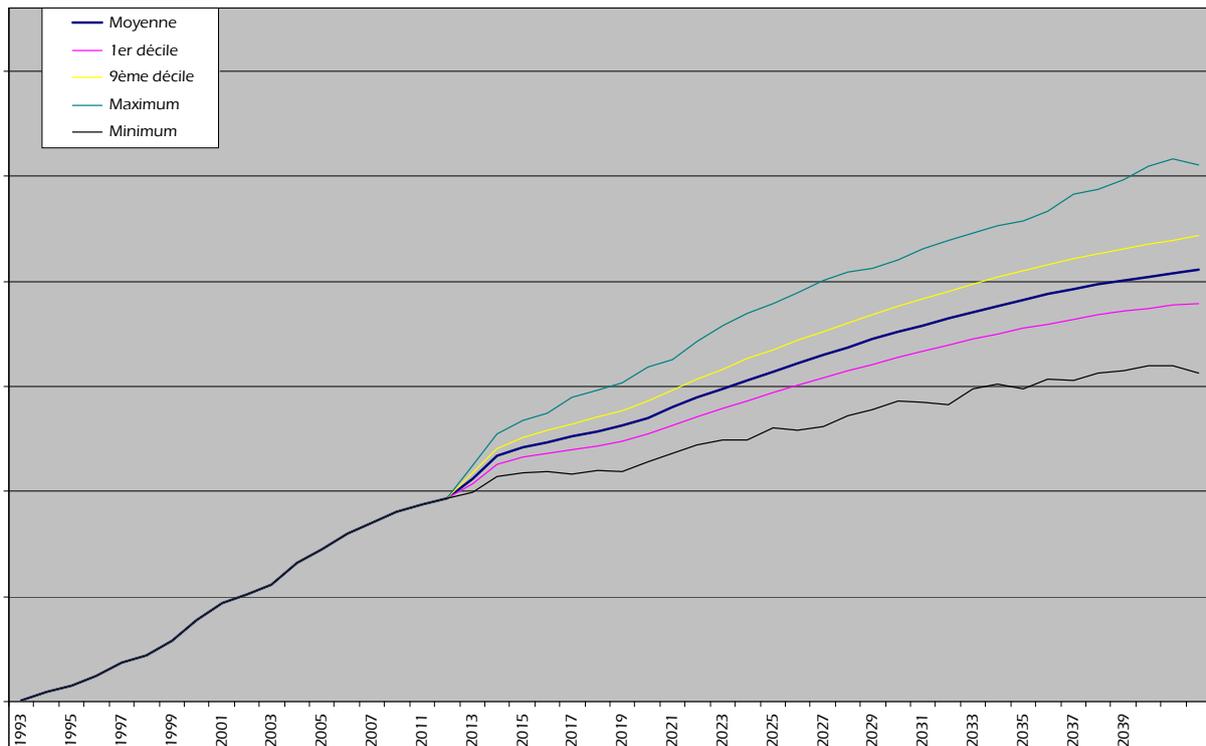
2. Projections à réglementation envisagée (n/2 entrées)

a. Scénario 1

Ce scénario prend en compte le changement réglementaire envisagé dès l'horizon 2011 en répliquant, pour les hommes, les comportements observés des femmes quant à leur demande d'avantage par mariage sous la règle actuelle.

Les résultats sont présentés graphiquement ci-après :

Projection de la population, sous la réglementation envisagée Avec correction de la population de base et l'hypothèse d'un comportement des hommes aligné sur celui des femmes



Les résultats sont résumés dans le tableau ci-après. Ils sont exprimés en pourcentage de croissance par rapport à la population de base connue en 2010 :

	2020	2030	2040
Projection moyenne	11,46%	20,54%	26,09%
Nombre minimum	6,06%	10,72%	14,29%
Nombre maximum	17,95%	29,45%	38,08%
1er décile	9,35%	17,48%	22,27%
9ème décile	13,57%	23,65%	29,92%

Ainsi, d'après ces résultats, la population pourrait vraisemblablement s'accroître, à l'horizon 2020, de 9.35% à 13.57% (en enlevant les simulations extrêmes), avec une moyenne probable de croissance égal à 11.46%.

De même, la population pourrait augmenter de 17.48% à 23.65% à horizon 2030 et de 22.27% à 29.92% à horizon 2040.

Par rapport à la réglementation actuelle, la population augmenterait en moyenne de :

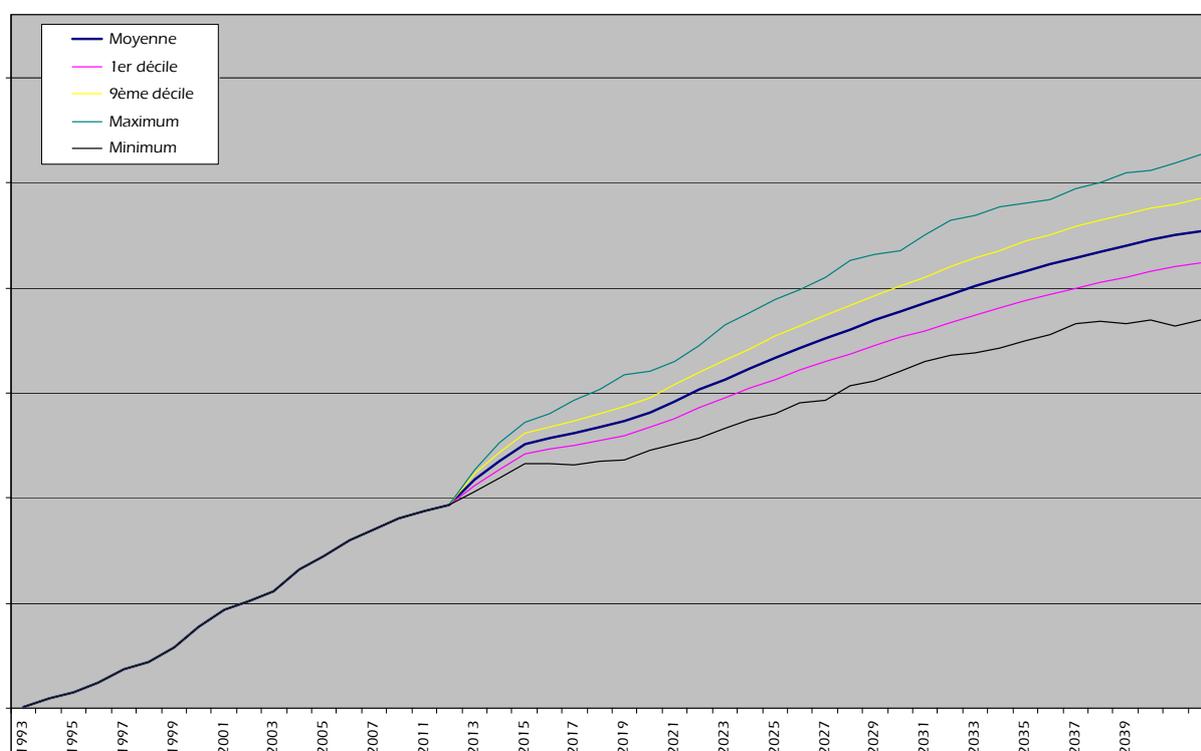
- 3.29% en 2020
- 5.20% en 2030
- 6.87% en 2040

b. Scénario 2

Ce scénario prend en compte le changement réglementaire envisagé dès l'horizon 2011 avec pour hypothèse que 100% des « étrangers » auront demandé l'avantage passé un délai de 3 ans suivant les 10 ans de mariage requis.

Les résultats sont présentés graphiquement ci-après :

Projection de la population, sous la réglementation envisagée Avec correction de la population de base et l'hypothèse de 100 % de demande d'avantage



Les résultats sont résumés dans le tableau ci-après. Ils sont exprimés en pourcentage de croissance par rapport à la population de base connue en 2010 :

	2020	2030	2040
Projection moyenne	13,13%	24,04%	31,31%
Nombre minimum	7,67%	17,01%	21,04%
Nombre maximum	18,12%	32,43%	40,01%
1er décile	11,03%	20,78%	27,65%
9ème décile	15,16%	27,19%	35,03%

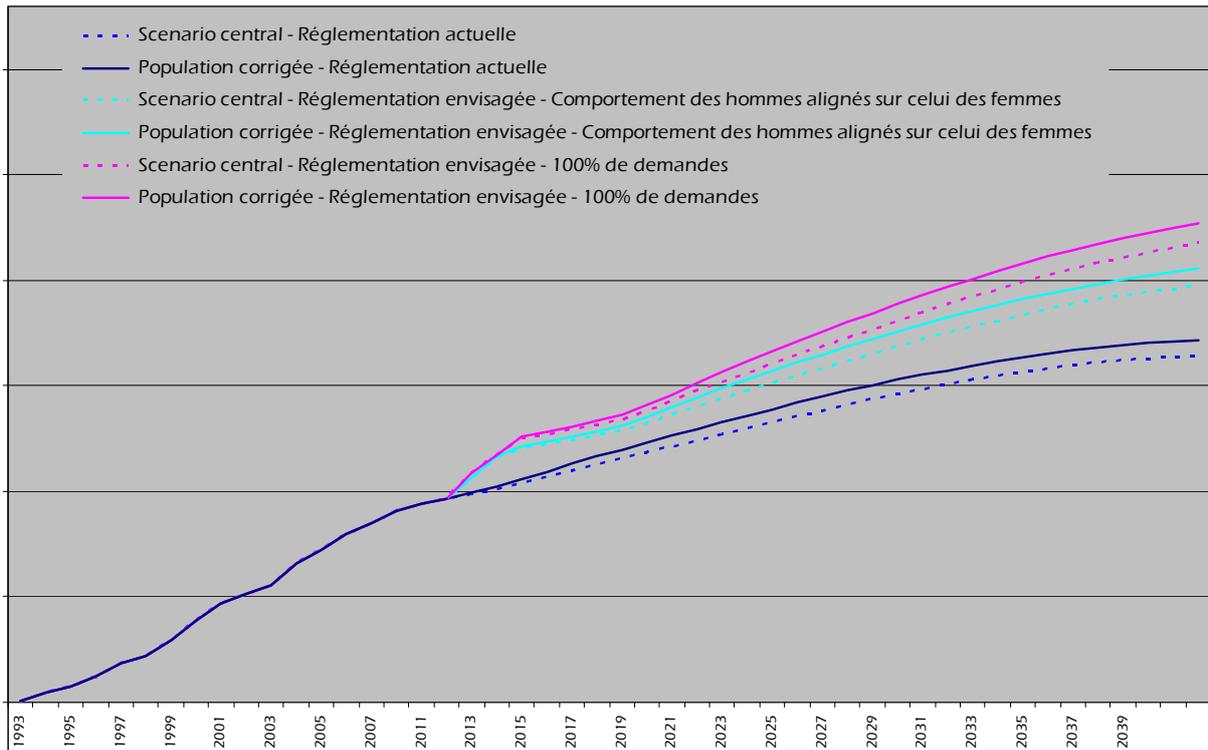
Ainsi, d'après ces résultats, la population pourrait vraisemblablement croître, à l'horizon 2020, entre 11.03% et 15.16% (en enlevant les simulations extrêmes), avec une moyenne de croissance probable à 13.13%.

De même, la population pourrait vraisemblablement augmenter de 20.78% à 27.19% à horizon 2030 et entre 27.65% à 35.03% personnes à horizon 2040.

Par rapport à la réglementation actuelle, la population augmenterait en moyenne de :

- 4.85% en 2020
- 8.25% en 2030
- 11.30% en 2040

Comparaison des résultats de ces simulations dans lesquelles la population de base a été corrigée avec celles du scénario de référence (moyennes des projections réalisées dans chacun des scénarios) :



Le nombre d'individus obtenu aux différents horizons et scénarios est légèrement plus élevé dans le cas où la base de données a été corrigée que sans correction. Cela provient du fait que nous avons dupliqué une population cible qui a plus d'enfants que la moyenne.

B. Projections sans prise en compte des entrées

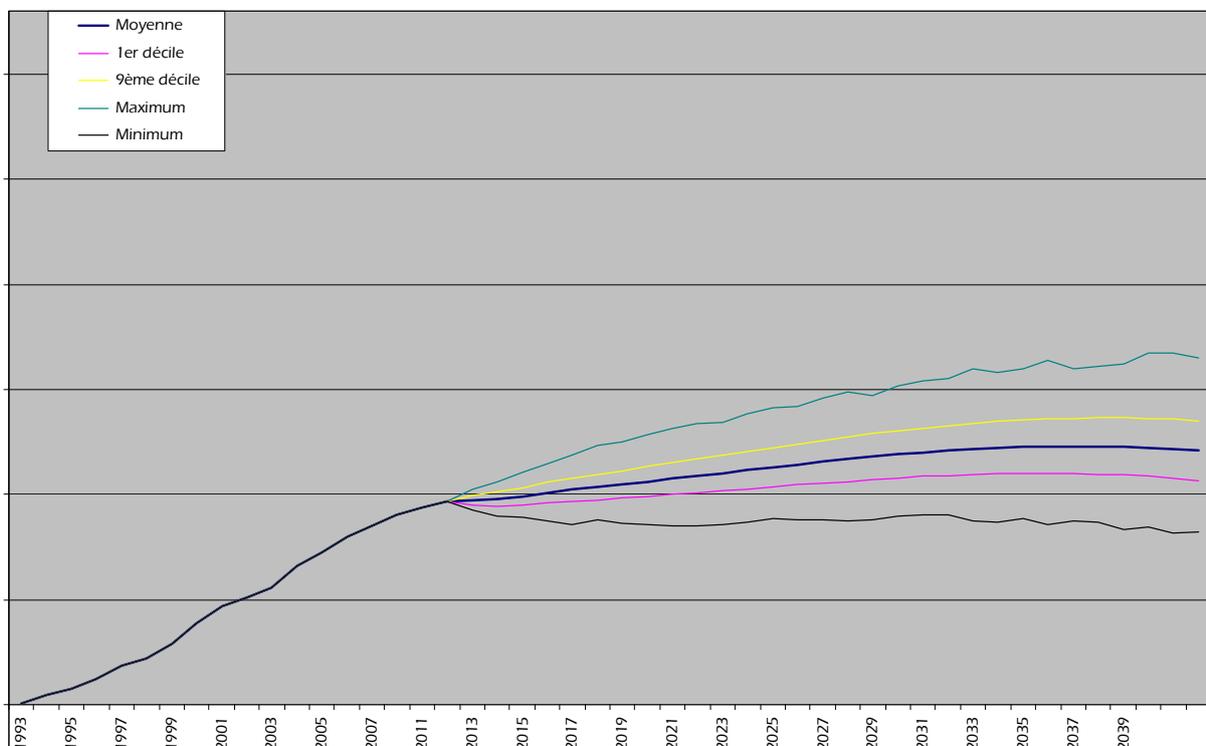
Le nombre d'entrée retenu dans les projections centrales est moitié moindre que la moyenne historique des dix dernières années. En fait, ce nombre d'entrée est extrêmement variable d'une année sur l'autre. Ainsi, par le passé, il est observé un rapport de 7 entre le nombre maximum d'entrées et le nombre minimum.

Aussi avons-nous réalisé deux études de sensibilité au nombre d'entrée dans le futur. La première, présentée ci-dessous, n'anticipe aucune entrée. La seconde au contraire, présentée en partie C, prévoit la moyenne historique du nombre d'entrées chaque année (soit le double que dans le scénario central).

1. Projections à réglementation inchangée

Les résultats sont présentés graphiquement ci-après :

Projection de la population, sous la réglementation actuelle Sans prise en compte des entrées



Les résultats sont résumés dans le tableau ci-après. Ils sont exprimés en pourcentage de croissance par rapport à la population de base connue en 2010 :

	2020	2030	2040
Projection moyenne	2,91%	5,77%	5,79%
Nombre minimum	-2,78%	-1,57%	-3,56%
Nombre maximum	8,81%	14,08%	16,37%
1er décile	1,00%	2,95%	2,37%
9ème décile	4,84%	8,63%	9,19%

Ainsi, d'après ces résultats, la population pourrait vraisemblablement s'accroître, à horizon 2020, de 1.00% à 4.84% (en enlevant les simulations extrêmes), avec une moyenne de croissance probable égale à 2.91%.

De même, la population pourrait avoir une croissance comprise de 2.95% à 8.63% à horizon 2030 et de 2.37% à 9.19% à horizon 2040.

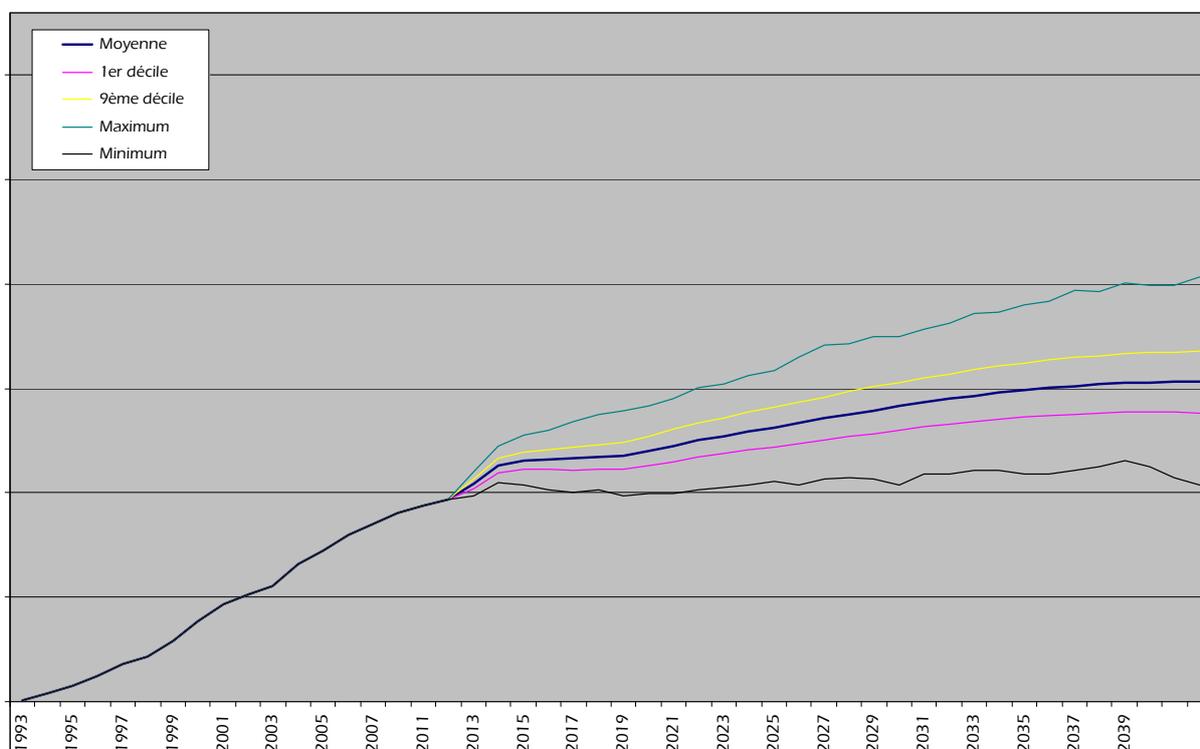
2. Projections à réglementation envisagée

a. Scénario 1

Ce scénario prend en compte le changement réglementaire envisagé dès l'horizon 2011 en répliquant, pour les hommes, les comportements observés des femmes quant à leur demande d'avantage par mariage sous la règle actuelle.

Les résultats sont présentés graphiquement ci-après :

Projection de la population, sous la réglementation envisagée Sans prise en compte des entrées et l'hypothèse d'un comportement des hommes aligné sur celui des femmes



Les résultats sont résumés dans le tableau ci-après. Ils sont exprimés en pourcentage de croissance par rapport à la population de base connue en 2010 :

	2020	2030	2040
Projection moyenne	6,81%	11,57%	13,53%
Nombre minimum	1,13%	2,90%	1,66%
Nombre maximum	12,94%	20,23%	25,58%
1er décile	4,86%	8,69%	10,00%
9ème décile	8,77%	14,48%	17,04%

Ainsi, d'après ces résultats, la population pourrait vraisemblablement s'accroître, à horizon 2020, de 4.86% à 8.77% (en enlevant les simulations extrêmes), avec une moyenne de croissance probable égale à 6.81%.

De même, la population pourrait avoir une croissance comprise entre 8.69% et 14.48% à horizon 2030 et entre 10.00% et 17.04% à horizon 2040.

Par rapport à la réglementation actuelle, la population augmenterait en moyenne de :

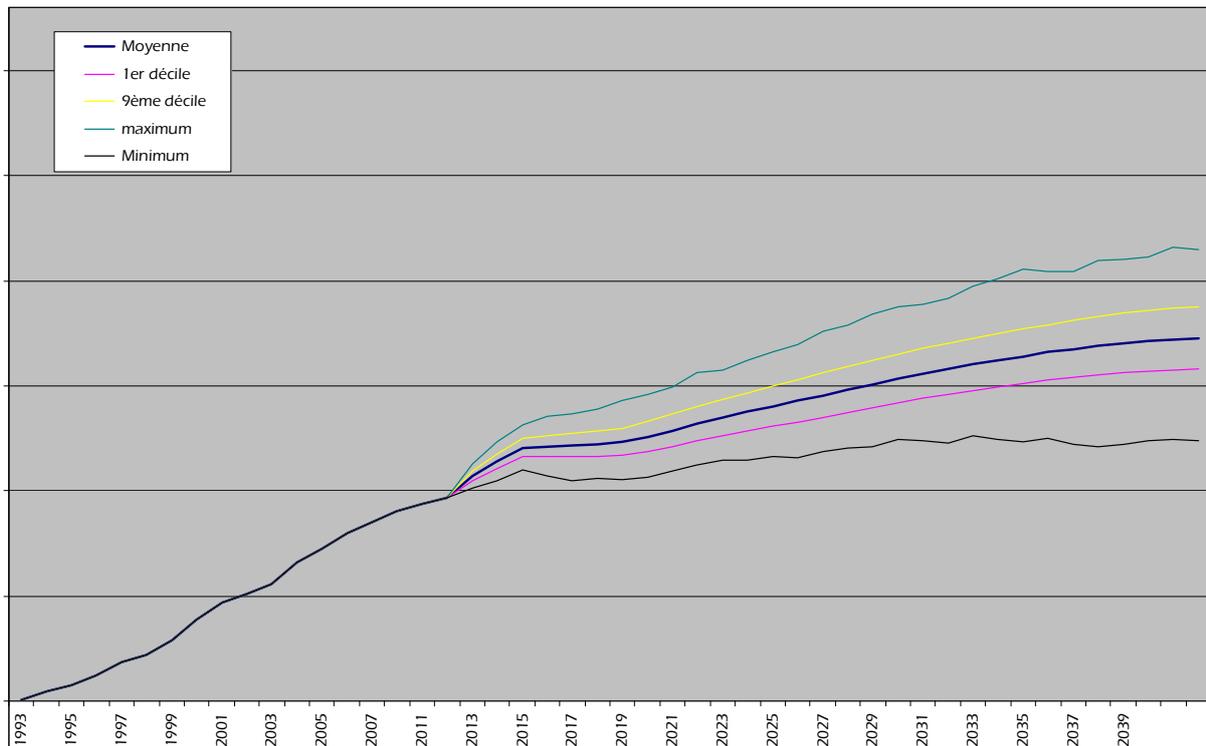
- 3.79% en 2020
- 5.48% en 2030
- 7.30% en 2040

b. Scénario 2

Ce scénario prend en compte le changement réglementaire envisagé dès l'horizon 2011 avec pour hypothèse que 100% des « étrangers » auront demandé l'avantage passé un délai de 3 ans suivant les 10 ans de mariage requis.

Les résultats sont présentés graphiquement ci-après :

Projection de la population, sous la réglementation envisagée Sans prise en compte des entrées et l'hypothèse de 100 % de demandes d'avantage



Les résultats sont résumés dans le tableau ci-après. Ils sont exprimés en pourcentage de croissance par rapport à la population de base connue en 2010 :

	2020	2030	2040
Projection moyenne	8,48%	14,69%	18,19%
Nombre minimum	3,80%	6,23%	6,50%
Nombre maximum	14,25%	22,79%	28,34%
1er décile	6,53%	11,82%	14,70%
9ème décile	10,46%	17,65%	21,78%

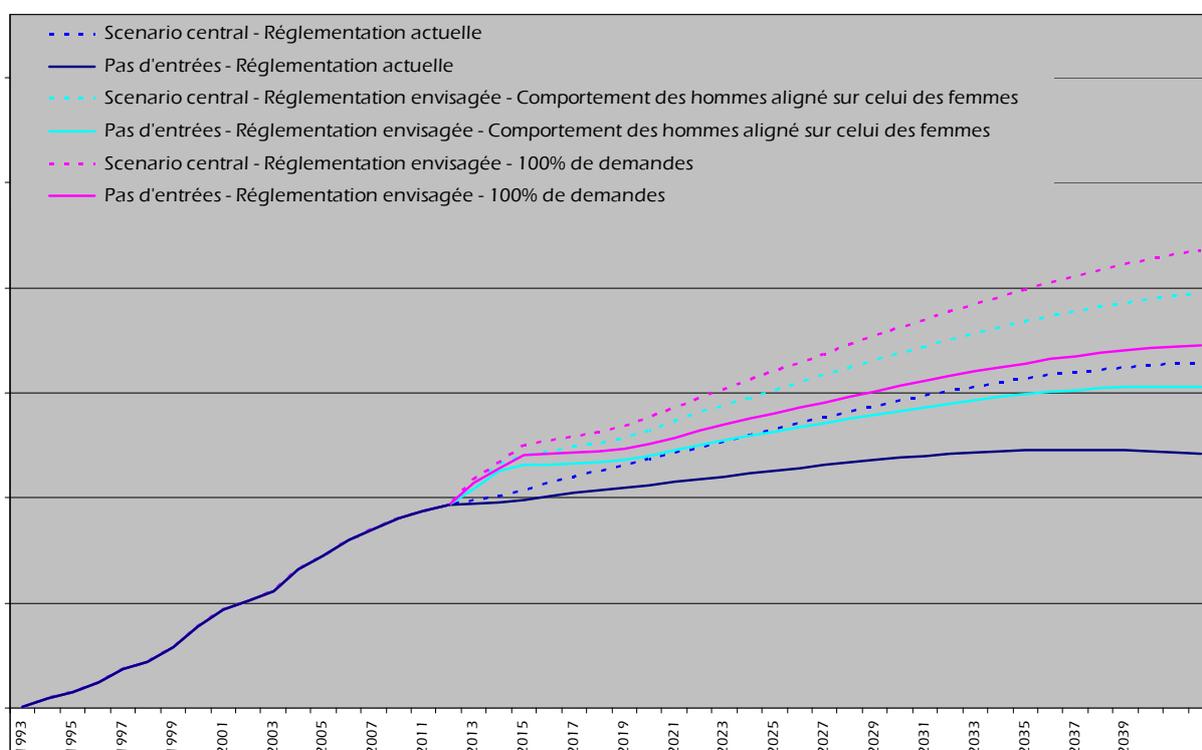
Ainsi, d'après ces résultats, la population pourrait vraisemblablement s'accroître, à horizon 2020, de 86.53% à 10.46% (en enlevant les simulations extrêmes), avec une moyenne de croissance probable égale à 8.48%.

De même, la population pourrait avoir une croissance comprise entre 11.82% et 17.65% à horizon 2030 et entre 14.70 et 21.78% à horizon 2040.

Par rapport à la réglementation actuelle, la population augmenterait en moyenne de :

- 5.42% en 2020
- 8.43% en 2030
- 11.73% en 2040

Comparaison des résultats de ces simulations dans lesquelles il n'y a aucune entrée avec celles du scénario de référence (moyennes des projections réalisées dans chacun des scénarios) :



L'hypothèse qu'il n'y aurait plus aucunes entrées dans le futur réduit fortement la perspective d'évolution future de la population. En revanche elle est pratiquement sans effet sur l'impact d'un changement de réglementation.

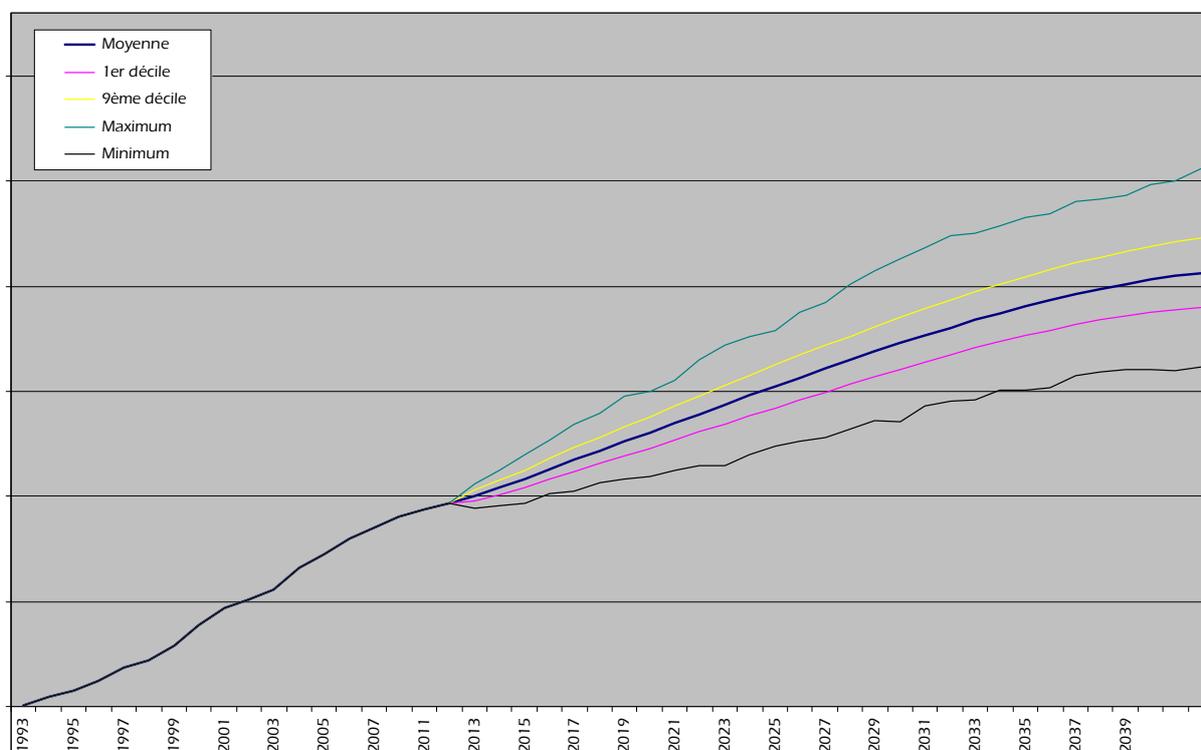
C. Projections en considérant la moyenne historique des entrées par an

Dans ce scenario, nous avons projeté la population avec une hypothèse de n entrées en moyenne chaque année, n représentant la moyenne du nombre d'entrée annuelle sur 10 ans (soit le double que dans le scenario central).

1. Projections à réglementation inchangée

Les résultats sont présentés graphiquement ci-après :

Projection de la population, sous la réglementation actuelle n entrées par an en moyenne



Les résultats sont résumés dans le tableau ci-après. Ils sont exprimés en pourcentage de croissance par rapport à la population de base connue en 2010 :

	2020	2030	2040
Projection moyenne	10,18%	20,05%	26,28%
Nombre minimum	4,35%	11,64%	15,49%
Nombre maximum	16,36%	30,51%	38,16%
1er décile	8,12%	16,91%	22,40%
9ème décile	12,25%	23,21%	30,28%

Ainsi, d'après ces résultats, la population pourrait vraisemblablement s'accroître, à horizon 2020, de 8.12% à 12.25% (en enlevant les simulations extrêmes), avec une moyenne de croissance probable à 10.18%.

De même, la population pourrait vraisemblablement avoir une croissance comprise entre 16.91% et 23.21% à horizon 2030 et entre 22.40% et 30.28% à horizon 2040.

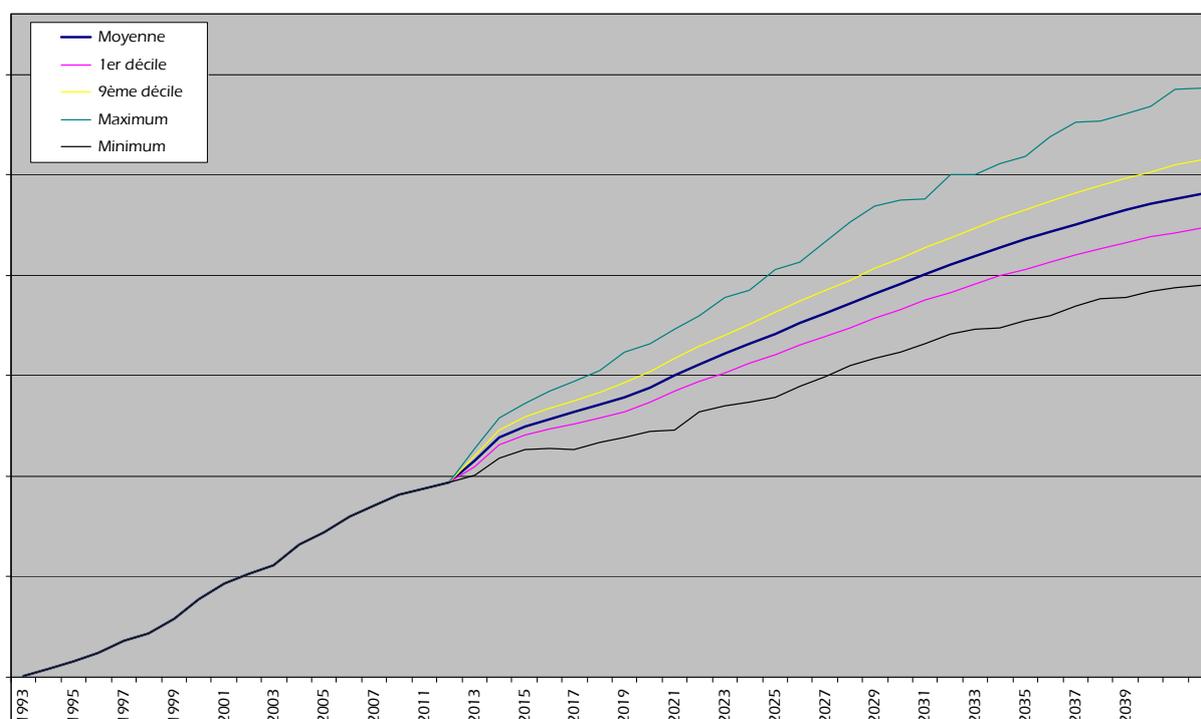
2. Projections à réglementation envisagée

a. Scénario 1

Ce scénario prend en compte le changement réglementaire envisagé dès l'horizon 2011 en répliquant, pour les hommes, les comportements observés des femmes quant à leur demande d'avantage par mariage sous la règle actuelle.

Les résultats sont présentés graphiquement ci-après :

Projection de la population, sous la réglementation envisagée n entrées par an en moyenne et l'hypothèse d'un comportement des hommes aligné sur celui des femmes



Les résultats sont résumés dans le tableau ci-après. Ils sont exprimés en pourcentage de croissance par rapport à la population de base connue en 2010 :

	2020	2030	2040
Projection moyenne	14,12%	25,99%	34,51%
Nombre minimum	8,41%	17,72%	23,56%
Nombre maximum	19,94%	36,81%	47,11%
1er décile	12,06%	22,77%	30,44%
9ème décile	16,23%	29,27%	38,54%

Ainsi, d'après ces résultats, la population pourrait vraisemblablement s'accroître, à horizon 2020, de 12.06% à 16.23% (en enlevant les simulations extrêmes), avec une moyenne de croissance probable égale à 14.12%.

De même, la population pourrait présenter une croissance comprise entre 22.77% et 29.27% à horizon 2030 et entre 30.44% et 38.54% à horizon 2040.

Par rapport à la réglementation actuelle, la population augmenterait en moyenne de :

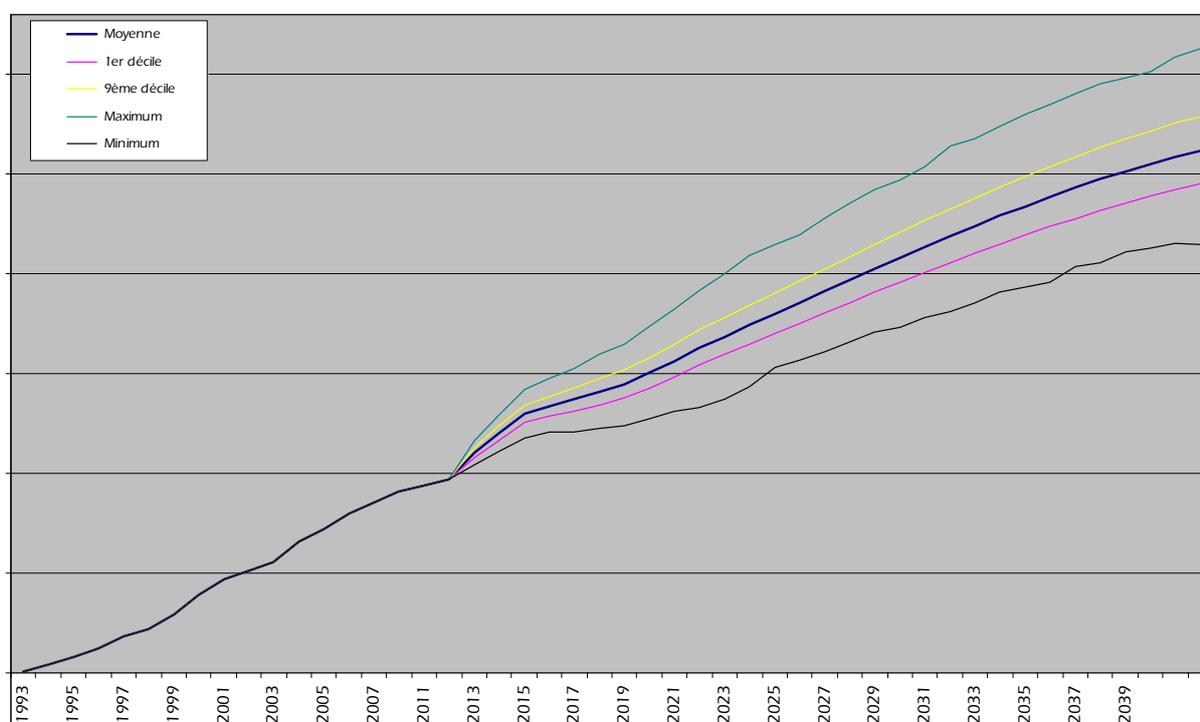
- 3.58% en 2020
- 4.94% en 2030
- 6.51% en 2040

b. Scénario 2

Ce scénario prend en compte le changement réglementaire envisagé dès l'horizon 2011 avec pour hypothèse que 100% des « étrangers » auront demandé l'avantage passé un délai de 3 ans suivant les 10 ans de mariage requis.

Les résultats sont présentés graphiquement ci-après :

Projection de la population, sous la réglementation envisagée n entrées par an en moyenne et l'hypothèse de 100 % de demandes d'avantage



Les résultats sont résumés dans le tableau ci-après. Ils sont exprimés en pourcentage de croissance par rapport à la population de base connue en 2010 :

	2020	2030	2040
Projection moyenne	15,85%	29,32%	39,62%
Nombre minimum	8,70%	20,27%	28,29%
Nombre maximum	22,69%	40,21%	51,81%
1er décile	13,76%	26,11%	35,61%
9ème décile	17,99%	32,59%	43,75%

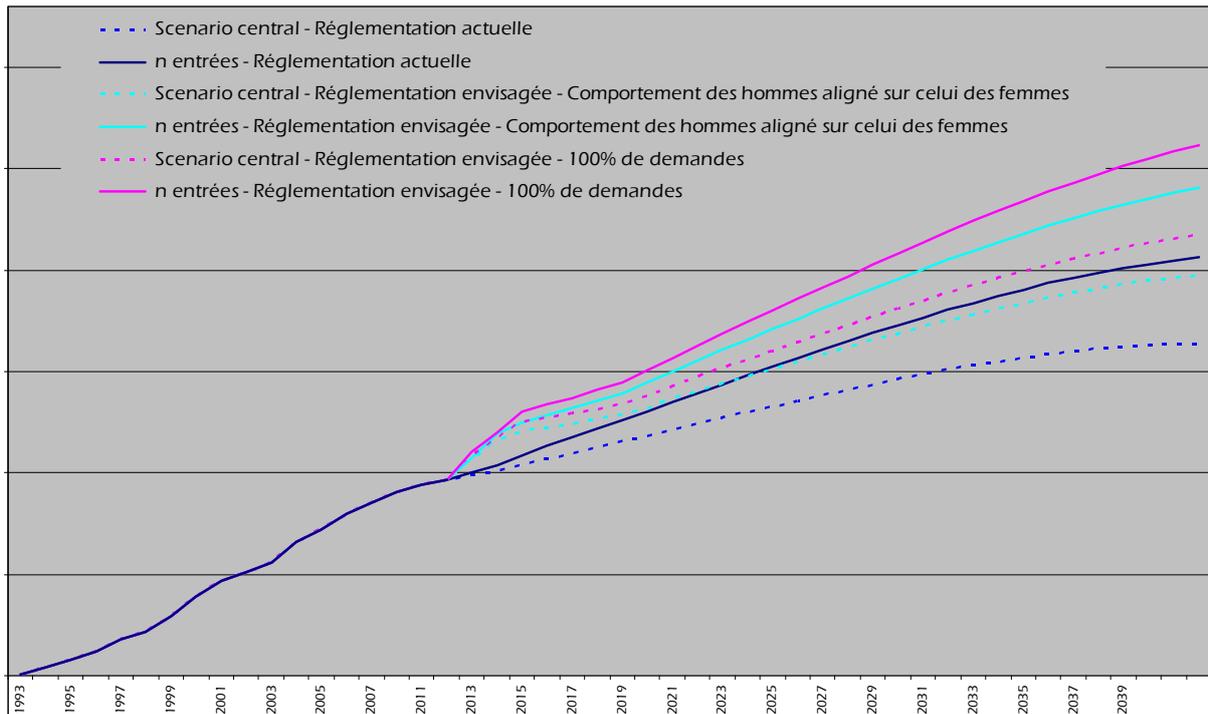
Ainsi, d'après ces résultats, la population pourrait vraisemblablement s'accroître, à horizon 2020, de 13.76% à 17.99% (en enlevant les simulations extrêmes), avec une moyenne de croissance probable à 15.85%.

De même, la population pourrait être avoir une croissance comprise entre 26.11% et 32.59% à horizon 2030 et entre 35.61% et 43.75% à horizon 2040.

Par rapport à la réglementation actuelle, la population augmenterait en moyenne de :

- 5.15% en 2020
- 7.70% en 2030
- 10.55% en 2040

Comparaison des résultats de ces simulations dans lesquelles il n'y a n entrées par an avec celles du scénario de référence (moyennes des projections réalisées dans chacun des scénarios) :



L'hypothèse qu'il y aurait n entrées en moyenne tous les ans dans le futur augmente la perspective d'évolution future de la population. En revanche elle est pratiquement sans effet sur l'impact d'un changement de règle.

Bibliographie

- Attal-toubert .K et Vanderschelden .M** (2009), « Mise en œuvre de la microsimulation : l'exemple du modèle de projections d'effectifs de médecins », Journée de méthodologie statistique, *DREES*.
- Baroni E. et Richiardi M.** (2007), « Orcutt's Vision, 50 years on », *LABOR*.
- Caisse Nationale d'Assurance Vieillesse** (2008), « *Retraite et Société* » n°54.
- Caisse Nationale d'Assurance Vieillesse** (2005), « *Retraite et Société* » n°48.
- Caisse Nationale d'Assurance Vieillesse** (2009), « Présentation du modèle PRISME », *Séminaire INED 6 janvier 2009*.
- Conseil d'orientation des retraites** (2005), « Fonds de réserve pour les retraites et modèles de projection », *Modèle de la CNAV*.
- De Gans H-A.** (2002) « Loi mathématique ou conjecture spéculative ? Un débat des années 1920 sur la méthodologie des projections démographiques », *Population*, p.91-119.
- Equipe DEMOSIM** (2010), « Projections de la diversité de la population canadienne », *Statistiques Canada*.
- Mahieu R. et Sédillot B.** (2000), « Microsimulations of the retirement decision: a supply-side approach », *INSEE*.
- Ministère de l'Economie, des Finances et de l'Industrie** (2004), « Microsimulation : l'expérience française », *Economie et prévision*.
- Privat A-G** (2005), « Evaluation de l'impact des réformes de 1993 et 2003 à l'aide du modèle de microsimulation ARTEMIS », *thèse présentée pour l'obtention du doctorat en science économique*.
- Robert-Bobée .I** (novembre 2001), « Modelling demographic behaviours in the French microsimulation model Destinie: An analysis of future change in completed fertility », *Série des documents de travail de la Direction des Etudes et Synthèses Économiques*.
- Robert-Bobée .I** (2006), « Projections de population 2005-2050 pour la France métropolitaine : méthode et résultats », *Insee, division Enquêtes et Etudes Démographiques*.
- Spielauler .M,** « Qu'est ce qu'une microsimulation dynamique en sciences sociales? », *Statistiques Canada-Division de la modélisation*.
- Stock J-H et Wise A.** (1990), « Pensions, the Option Value of Work, and Retirement », *ECONOMETRICA vol.58 p.1151-1180*.

Van Imhoff .E et Post .W (1997), « Méthodes de microsimulation pour des projections de populations » *Population*, 4, p.889-932.