

Projections probabilistes bayésiennes de population pour la France

Bayesian Probabilistic Population Projections for France

Vianney Costemalle*

Résumé – Les projections de population sont un exercice régulier des instituts nationaux de statistique. En France, les dernières ont été produites en 2016 par l’Insee en utilisant une approche déterministe basée sur 27 scénarios différents. Dans cet article nous proposons une nouvelle approche en projetant de façon probabiliste la population et en utilisant de plus le paradigme bayésien, afin de quantifier l’incertitude sur le niveau des populations futures sans recourir à des scénarios. Selon la méthode des composantes, on projette de manière indépendante et par sexe et âge le taux de mortalité, le taux de fécondité et le solde migratoire. Ces trois composantes sont modélisées en tenant compte des données d’état civil (nombre de naissances et de décès) et des séries du solde migratoire. Les résultats montrent que la population de la France métropolitaine continuera de croître pour atteindre un niveau compris entre 66.1 millions et 77.2 millions d’habitants en 2070 avec une probabilité de 95 %.

Abstract – Population projections are performed regularly by national statistics institutes. In France, the most recent projections were produced by Insee in 2016 using a deterministic approach based on 27 different scenarios. In this article, we propose a new approach which combines probabilistic population projections and a greater use of the Bayesian paradigm in order to quantify the uncertainty of future population levels without resorting to scenarios. Using the components method, the mortality rate, fertility rate and net migration are projected independently by sex and age. These three components are modelled, taking account of registry data (number of births and deaths) and net migration data series. The results reveal that the population of metropolitan France will continue to grow, reaching a level of between 66.1 million and 77.2 million inhabitants in 2070, with a probability of 95%.

Codes JEL / JEL Classification : C11, C53, J11, J13, F22

Mots clés : projections probabilistes, inférence bayésienne, séries temporelles, population, mortalité, fécondité, migrations

Keywords: probabilistic projections, Bayesian inference, time series, population, mortality, fertility, migration

* Insee au moment de la rédaction de cet article (vianney.costemalle@sante.gouv.fr)

Remerciements – L’auteur tient à remercier Julian Arbel, Junni Zhang, John Bryant, Marie Reynaud, Isabelle Robert-Bobée, Nathalie Blanpain, Guillemette Buisson, Vanessa Bellami, Catherine Beaumel ainsi que deux rapporteurs anonymes.

Reçu en novembre 2017, accepté en juillet 2018.

Citation: Costemalle, V. (2020). Bayesian Probabilistic Population Projections for France. *Economie et Statistique / Economics and Statistics*, 520-521, 29–47. <https://doi.org/10.24187/ecostat.2020.520d.2031>

Les projections de population sont un exercice régulier des instituts de statistique à travers le monde ainsi que de certaines organisations internationales comme l'Organisation des Nations Unies (ONU) qui tous les deux ou trois ans depuis 1951 publie les *World Population Prospects* (ONU, 2017). L'intérêt des projections de population est multiple et ses utilisateurs sont nombreux. Elles servent avant tout à prévoir, sous certaines hypothèses, ce que pourrait être la population d'une région, d'un pays ou du monde entier, en nombre d'habitants ainsi qu'en structure. À court ou moyen terme, ces projections sont à la base de la planification économique et sociale comme par exemple le financement des retraites (COR, 2017) ou la construction d'infrastructures publiques. Elles sont aussi un élément essentiel dans certains autres exercices comme les projections économiques, climatiques ou environnementales.

En ce qui concerne la France, les dernières projections officielles datent de 2016 (Blanpain & Buisson, 2016a et 2016b). Elles indiquent quelle sera la population en 2070 si les tendances passées se poursuivent, avec différentes variantes sur ces hypothèses (voir Blanpain dans ce numéro). Le détail des projections par régions, et notamment celles portant sur la France métropolitaine, n'est disponible que pour la période 2013-2050. L'objectif de cet article est d'explorer une nouvelle méthode pour projeter la population de la France : les projections probabilistes. L'approche proposée est dite probabiliste car elle permet de quantifier l'incertitude sur le niveau de la population future ; elle se distingue en cela de l'approche traditionnelle qui est un ensemble de projections déterministes basées sur différents scénarios. La différence fondamentale entre ces deux approches ne tient pas tellement aux résultats eux-mêmes mais plutôt à la façon dont ils sont interprétés et utilisés.

Les projections probabilistes reposent sur des modèles statistiques, la plupart du temps paramétriques. L'incertitude sur certaines composantes de la population peut être captée par des termes d'erreurs, comme dans le cas des séries temporelles, mais elle peut aussi provenir d'une inférence bayésienne des paramètres du modèle. Tout l'objectif est de quantifier l'incertitude sur la population future. Pour cela, on peut utiliser l'approche stochastique, l'approche bayésienne, ou même une combinaison des deux. Dans cet article, nous utilisons des modèles stochastiques avec inférence bayésienne des paramètres.

Dans une tribune libre du *Journal of Official Statistics*, un ensemble de démographes et

d'universitaires de différents pays mentionnent les apports et les défis des projections probabilistes en démographie et appellent à plus de recherche et de pratique dans ce domaine de la part des instituts de statistique (Bijak *et al.*, 2015). Ils soulignent le fait que les projections probabilistes ont déjà été développées et utilisées avec succès dans d'autres disciplines comme la météorologie, la climatologie ou l'aviation. Les statistiques bayésiennes mettent également du temps à pénétrer le champ de la démographie. Bien que le théorème de Bayes ait été établi il y a plus de 250 ans, ce n'est que récemment, avec l'apparition des algorithmes MCMC (Markov Chains Monte-Carlo) à partir des années 1980 et avec l'explosion de la puissance de calcul des ordinateurs, que l'inférence bayésienne est mise en œuvre (Bijak & Bryant, 2016).

Certains instituts de statistique ont déjà adopté la démarche visant à produire des projections démographiques probabilistes pour leurs statistiques officielles. C'est le cas en particulier des Pays-Bas et de la Nouvelle-Zélande. Les Pays-Bas ont commencé à produire des projections probabilistes basées sur des méthodes stochastiques dès 1998. La Nouvelle-Zélande communique également depuis 2012 des résultats probabilistes de projections de population (MacPherson, 2016 ; Dunstan & Ball, 2016). Enfin l'ONU, qui réalise des projections pour l'ensemble des pays, est passée d'une méthode déterministe à une méthode probabiliste en 2014 (Costemalle, 2015). De plus certaines composantes de ces projections sont basées sur l'inférence bayésienne.

La très grande majorité des projections de population repose sur la méthode des composantes, qui consiste à projeter séparément les trois composantes essentielles de la dynamique des populations à savoir la fécondité, la mortalité et les migrations. La population à une période donnée est décomposée par sexe et catégories d'âge et elle est égale à la population de la période précédente à laquelle on ajoute les naissances et les immigrants et de laquelle on retire les décès et les émigrants. De cette façon on peut faire évoluer, période par période, la population et sa structure par sexe et catégories d'âge. Pour cela, il faut à chaque période, déterminer le nombre de naissances par sexe ainsi que le nombre de décès et le solde migratoire par sexe et catégorie d'âge. En ce qui concerne les naissances et les décès, les méthodes les plus répandues reposent sur la projection des taux de fécondité et des taux de mortalité. Mais les projections probabilistes de population restent un domaine de recherche actif : il n'existe pas de méthode unique, il y a

au contraire presque autant d'approches que de données qui diffèrent d'un pays à un autre.

Dans cet article, nous soulignons dans une première section les différences essentielles entre les projections déterministes et les projections probabilistes, puis nous présentons quelques-unes des différentes approches qui ont été développées en démographie en ce qui concerne les projections probabilistes de population. La deuxième section est consacrée à la description des données françaises de mortalité, de fécondité et de solde migratoire, la troisième à la présentation et à la validation des modélisations retenues pour chacune des trois composantes. Enfin nous présentons les résultats des projections probabilistes ainsi obtenues pour la France, avant de discuter les hypothèses des modèles.

1. Projections déterministes, projections probabilistes et développements en démographie

Prévoir l'avenir est un exercice difficile et de nombreuses méthodes se sont développées au cours des années. Les méthodes les plus récentes et sophistiquées se basent sur des modèles mathématiques qui tentent de détecter certains motifs ou invariants dans les données et de prolonger les tendances observées, tout en respectant certaines contraintes que l'on peut s'imposer. Les projections déterministes et probabilistes font toutes deux appel à un certain degré de modélisation des données observées, elles ne diffèrent que sur la nature des prévisions.

1.1. Approches déterministe et probabiliste : différentes façons d'aborder l'avenir

Dans le premier cas, ce que l'on cherche à projeter dépend de façon déterministe de certains paramètres. Le choix de ces paramètres constitue une hypothèse que l'on nomme aussi scénario. On se donne alors un scénario d'évolution de ces paramètres, que l'on juge le plus probable au vu des connaissances accumulées, des avis des experts et de l'intuition. Un scénario donné correspond à une et une seule projection possible, et le rapport entre les deux est déterministe. Dans le cas où le scénario se réaliserait, la projection serait certaine. Les projections déterministes répondent donc à la question : « Que se passerait-il dans l'avenir dans le cas de l'avènement d'un tel scénario ? ». On peut ainsi formuler des scénarios extrêmes pour voir comment se comporterait alors le futur dans le cas de leur réalisation. Les projections déterministes sont donc un formidable outil pour

explorer l'avenir à partir de scénarios préétablis. Toute l'incertitude de la projection repose alors sur la réalisation du scénario. On formule des scénarios possibles, mais on n'est pas en mesure de savoir avec quel degré de probabilité ils pourront se réaliser. On peut même affirmer que la probabilité de leur réalisation est nulle (si les grandeurs sont continues) ou très faible (si les grandeurs sont discrètes). Le degré de probabilité est estimé de façon intuitive et se reflète dans les termes utilisés pour décrire ces scénarios : on parle de scénario « central », pour le scénario considéré comme le plus plausible compte tenu des connaissances actuelles, et de scénarios « extrêmes ».

Au contraire, les projections probabilistes sont basées sur des modèles qui essaient de tenir compte de l'incertitude résultant de l'ignorance de certains aspects des projections. Ces modèles reposent sur des hypothèses qui sont le fruit de jugements d'experts et des intuitions. Les hypothèses sous-jacentes des modèles dans les projections probabilistes sont l'équivalent des scénarios dans les projections déterministes. L'avantage des projections probabilistes est de quantifier l'incertitude à partir des évolutions observées par le passé et de la propager dans le futur afin d'avoir des intervalles de confiance des projections. Ainsi l'interprétation et l'utilisation des projections probabilistes diffèrent de celles des projections déterministes.

Les prévisions météorologiques utilisent par exemple depuis longtemps des projections probabilistes : on ne nous dit pas seulement s'il va pleuvoir ou non le lendemain, mais avec quelle probabilité il risque de pleuvoir (Raftery, 2014). Les événements futurs étant par nature incertains, indiquer la probabilité de leur réalisation, étant donné les connaissances actuelles, donne ainsi plus d'information qu'une projection déterministe basée sur un scénario. Les séries temporelles sont, en sciences économiques en particulier, un moyen de produire des projections probabilistes : dans le cas d'une marche aléatoire simple par exemple on sait que la variance augmente avec la racine carrée du temps.

En ajoutant des termes d'erreurs dans les modèles, on peut donc créer des projections probabilistes stochastiques. Une autre manière de quantifier l'incertitude est de s'appuyer sur le paradigme bayésien. Dans ce dernier, les paramètres des modèles sont considérés comme des variables aléatoires, au même titre que les termes d'erreurs dans les modèles stochastiques. L'inférence bayésienne consiste alors à estimer la distribution *a posteriori* de ces paramètres,

c'est-à-dire après l'observation des données. Cette distribution donne des valeurs possibles des paramètres et leur degré de probabilité. Elle diffère de la distribution *a priori* qui est la distribution donnée par le modélisateur et qui est censée refléter la connaissance du problème avant toute observation des données.

1.2. Les projections probabilistes en démographie : des modélisations en pratique très variées

On peut classer les techniques de projections de population en trois catégories (Booth, 2006). La première regroupe les méthodes basées sur l'extrapolation des tendances, qui cherchent à prolonger, de façon linéaire le plus souvent, les tendances détectées dans le passé. Elles se basent uniquement sur les données passées et ne cherchent pas à expliquer les mécanismes sous-jacents d'évolution. Elles se révèlent souvent efficaces. La deuxième façon de projeter la population est de se fixer des tendances de long terme. Ces méthodes sont basées sur le fait que l'on s'attend à ce que l'avenir se déroule d'une certaine manière. Cela peut être justifié par des avis d'experts, qui évaluent ce que l'on pourrait attendre pour le futur compte tenu de leurs connaissances actuelles, ou par des intentions de personnes, comme celles mesurées par les enquêtes d'intentions de fécondité (Régnier-Loilier & Vignoli, 2011). Enfin, la dernière catégorie de projection comprend les modèles structurels, qui essaient d'expliquer les mécanismes d'évolution avec des variables exogènes. Il faut alors projeter ces variables exogènes selon l'une des trois catégories de projection. Souvent, les approches proposées mêlent plusieurs de ces techniques et les techniques utilisées diffèrent selon les composantes (mortalité, fécondité et migration) que l'on souhaite projeter.

Une méthode classique de projection de la mortalité a été développée par Lee & Carter (1992) et consiste à décomposer l'évolution du logarithme des taux de mortalité en un effet de l'âge et un effet du temps, spécifique à chaque âge. L'effet temporel est ensuite considéré comme une série temporelle dont on estime les paramètres. Par calculs ou par la simulation, un très grand nombre de fois, des valeurs futures de cet effet temporel en faisant appel à la modélisation retenue, il est possible d'avoir une projection probabiliste. L'idée essentielle de cette approche est de capter dans les données les évolutions régulières et d'extrapoler ces régularités. La méthode de Lee-Carter a depuis été utilisée très fréquemment pour projeter la

mortalité, mais aussi pour projeter la fécondité et les migrations. Wiśniowski *et al.* (2015) en proposent une version plus étendue, en ajoutant un effet de génération, que l'on peut appliquer aux trois composantes des variations de la population. Ces auteurs proposent de plus de réaliser ces projections dans un cadre entièrement bayésien. Le modèle de Lee-Carter a également été généralisé par Hyndman & Ullah (2007) qui décomposent le logarithme des taux de mortalité, ou de fécondité, en composantes principales et qui prolongent les coefficients de chacune de ces composantes à l'aide de séries temporelles. Hyndman & Booth (2006) suggèrent de plus de réaliser une transformation de Box et Jenkins sur les taux étudiés afin de généraliser la transformation logarithmique. Cette approche est entièrement stochastique.

Tout l'intérêt des projections probabilistes est de pouvoir quantifier le degré de probabilité des projections futures. C'est ainsi qu'en 2001, Lutz *et al.* (2001) annoncent qu'il est probable que la population mondiale cesse de croître d'ici la fin du siècle. Plus précisément, leurs modèles stochastiques et leurs calculs prévoient que la population mondiale pourrait commencer à décroître d'ici la fin du siècle avec une probabilité de 85 %. L'ONU, qui publie régulièrement des projections de population, a commencé à utiliser une méthode probabiliste et bayésienne à partir de 2014. Les résultats donnent un aperçu différent de l'évolution de la population à long terme. Ils montrent en effet que la fin de la croissance mondiale de population est improbable d'ici à 2100 (Gerland *et al.*, 2014). La méthodologie est différente de celle de Lutz *et al.* (2001) : les grandeurs agrégées que sont l'espérance de vie à la naissance et l'indicateur conjoncturel de fécondité (ICF) sont projetées directement dans un premier temps, avant de décomposer ces indicateurs en taux de mortalité par sexe et âge et en taux de fécondité par âge. Pour projeter l'espérance de vie, le gain d'espérance de vie tous les cinq ans est modélisé par une double fonction logit, dépendant de l'espérance de vie actuelle et de nombreux paramètres. Ces paramètres sont estimés par inférence bayésienne, ce qui conduit à avoir une distribution *a posteriori* des gains d'espérance de vie, et donc une distribution *a posteriori* de l'espérance de vie elle-même à l'horizon 2100 (Raftery *et al.*, 2013). Ceci est l'exemple d'une projection probabiliste ne faisant pas intervenir de termes stochastiques, mais étant uniquement basée sur une modélisation paramétrique et une inférence bayésienne. L'ICF est quant à lui modélisé selon un processus d'évolution en trois phases : phase

de haute fécondité, phase de déclin rapide de la fécondité jusqu'en dessous du seuil de renouvellement des générations, et phase de stagnation de la fécondité avec une convergence à long terme vers un niveau à 2.1 enfants par femme (Alkema *et al.*, 2010).

Il apparaît donc que de nombreux modèles existent pour projeter chacune des trois composantes. Partant du principe qu'aucun modèle ne peut à lui seul rendre compte de l'ensemble des hypothèses possibles sur l'évolution de la mortalité, surtout lorsque ces hypothèses ne sont pas cohérentes entre elles, Kontis *et al.* (2017) se sont servi de 21 modèles différents de projections probabilistes, dont les résultats ont ensuite été pondérés selon la performance de chacun des modèles, pour au final déboucher sur une seule distribution de probabilité pour les indicateurs souhaités.

2. Les données pour la France

Afin de disposer de séries longues, on se restreint au champ de la France métropolitaine. On dispose ainsi pour les années 1962 à 2013, de façon détaillée par sexe et âge, du total de la population au 1^{er} janvier de chaque année, du solde migratoire annuel, du nombre de décès et du nombre de naissances selon l'âge de la mère¹. Nous choisissons le même horizon de projection que celui qui a été retenu pour les dernières projections officielles de la France (Blanpain & Buisson, 2016b). L'objectif est donc de projeter la population de 2014 à 2070. De 1962 jusqu'à 1998, les données ne sont pas détaillées par âge au-delà de 100 ans. À partir de 1999 elles sont ventilées en détail jusqu'à l'âge de 110 ans. Nous retenons des tranches d'âge d'un an, car les données sont disponibles, et créons une tranche d'âge supérieure, correspondant aux personnes de 100 ans et plus. Dans la suite de cette section, nous décrivons les données du solde migratoire, de la mortalité et de la fécondité, pour en dégager les invariants, les tendances, et aussi les irrégularités.

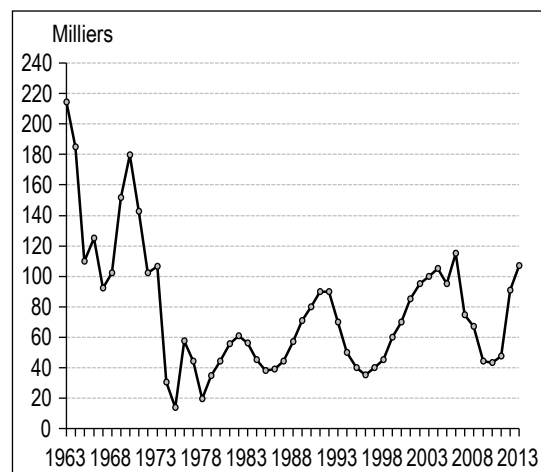
Le solde migratoire est, pour une année donnée, le nombre de personnes vivant hors de la France métropolitaine, quelle que soit leur nationalité, venant habiter en France, moins le nombre de personnes vivant en France métropolitaine et allant vivre en dehors du territoire. C'est sans doute la composante la plus difficile à mesurer, car bien que l'on puisse estimer les entrées à l'aide du recensement de la population (Brutel, 2014), on ne connaît pas les sorties. Le solde migratoire peut alors se déduire comme la différence entre l'évolution de la population

et le solde naturel. Contrairement à beaucoup d'autres pays européens, la France ne dispose pas de registre de la population et doit donc s'appuyer sur le recensement de la population pour estimer les flux migratoires. Jusqu'en 1999, le recensement n'ayant lieu que tous les 7-8 ans environ, l'évolution de la population entre deux années consécutives ne pouvait se calculer directement. En 1962, en raison du retour des français d'Algérie, le solde migratoire a été exceptionnellement très important, de l'ordre de 860 000 personnes ; depuis 1963 le solde migratoire atteint des niveaux toujours positifs, mais bien plus faibles : il vaut en moyenne 64 000 sur la période 1963-2013. Le solde reste stable en moyenne à partir des années 1990, même si de fortes fluctuations apparaissent (figure I), notamment dues aux différentes politiques menées, mais aussi au contexte économique et international. En moyenne sur la période 1990-2013, le solde migratoire est de 72 000 et de 79 000 sur les dix dernières années disponibles (2004-2013).

Pour décrire la mortalité, il faut rapporter le nombre de décès à la population à risque correspondante. Cette population se compte en personnes-années et prend en compte la durée totale de résidence de l'ensemble des personnes résidant en France. On l'estime en ajoutant à la population présente au 1^{er} janvier la moitié du solde migratoire. En rapportant le nombre de

1. 2013 est la dernière année pour laquelle toutes ces données étaient définitives lorsque les projections présentées dans cet article ont été réalisées en 2017. En particulier, le solde migratoire n'était à ce moment pas encore disponible pour l'année 2014. Nous n'avons pas utilisé les données provisoires, alors disponibles jusqu'en 2016 : révisées d'une année à l'autre avant de devenir définitives, elles sont en effet de nature différente des données définitives.

Figure I – Solde migratoire de 1963 à 2013



Source et champ : Insee, estimations de population et statistiques de l'état civil ; France métropolitaine.

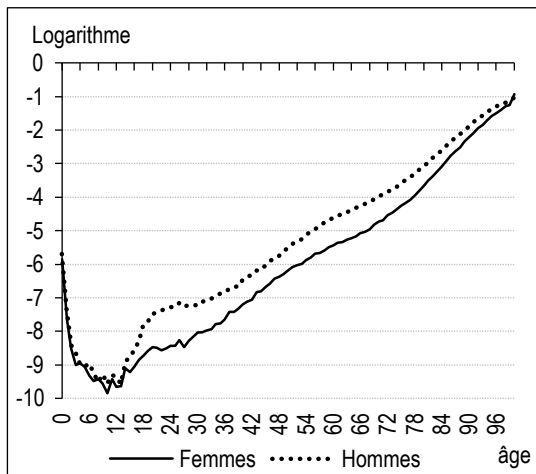
décès à cette population, on obtient les taux de mortalité, que l'on peut détailler par sexe, âge et année. Les taux de mortalité évoluent de manière quasi-exponentielle à partir de l'âge de 25 ans (figure II). Avant 25 ans, le profil est différent en raison de la mortalité infantile, plus élevée pour les nouveau-nés. Les taux de mortalité diminuent de la naissance jusqu'à l'âge de 10 ans environ, avant d'augmenter régulièrement. Vers 18 ans, la mortalité des hommes devient nettement plus élevée que celle des femmes, et l'écart persiste tout au long de la vie, avec une ampleur plus ou moins marquée selon l'âge.

Le logarithme des taux de mortalité, à âge et sexe fixés, diminue de façon quasi-linéaire avec le temps (figure III). Si c'est particulièrement vrai

pour les âges élevés, ça l'est moins pour les âges plus jeunes. Le logarithme du quotient de mortalité à 10 ans diminue de plus en plus vite par exemple. Au contraire, à 30 ans, le logarithme du quotient de mortalité a ralenti sa décroissance, jusqu'à stagner pour les hommes, du début des années 1980 au milieu des années 1990, date à laquelle la mortalité a brusquement diminué pour cet âge, et continue depuis lors à diminuer régulièrement et apparemment linéairement. Cette stagnation de la mortalité chez les jeunes adultes des années 1980 et 1990, alors que la tendance générale était à une diminution constante de la mortalité, est à relier à l'épidémie de Sida qui a touché la France au début des années 1980. De façon générale, les taux de mortalité baissant régulièrement, l'espérance de vie à la naissance augmente chaque année, et ce plus rapidement pour les hommes que pour les femmes (Blanpain, 2016), même s'il arrive parfois que l'espérance de vie baisse d'une année sur l'autre comme c'était le cas en 2015 pour des raisons conjoncturelles (Bellamy & Beaumel, 2016).

Depuis le début des années 1970, l'ICF² a fortement diminué : il est passé de 2.9 enfants par femme en 1964 à 1.8 enfant par femme en 1976 (figure IV). Depuis lors il s'est stabilisé autour d'une valeur moyenne de 1.85 enfant par femme. On observe toutefois une tendance à l'augmentation de l'ICF depuis le milieu des années 1990.

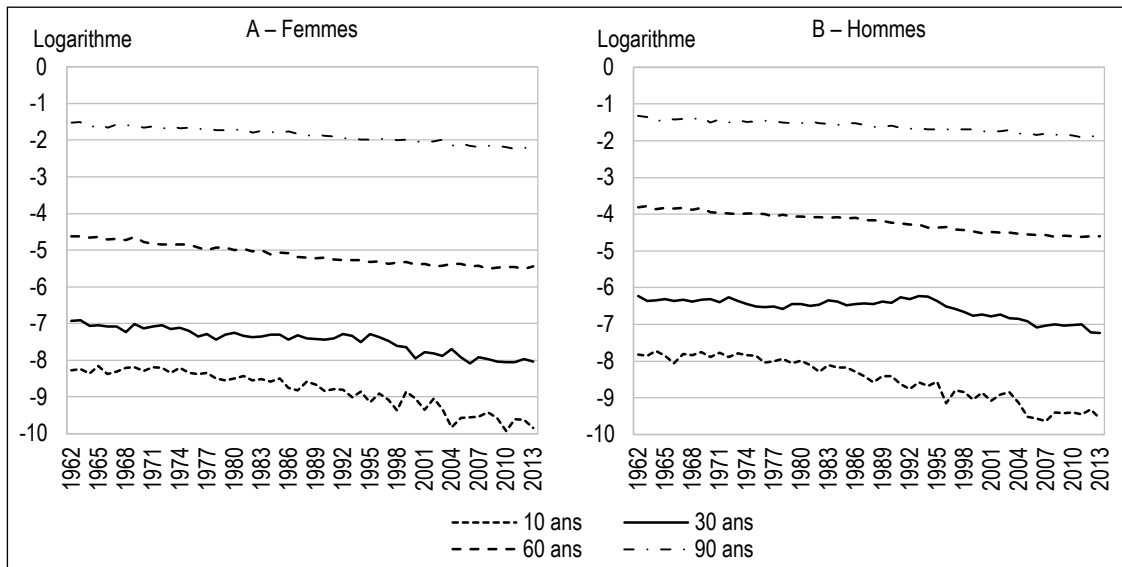
Figure II – Logarithme des taux de mortalité en 2013 selon le sexe et l'âge



Source et champ : Insee, estimations de population et statistiques de l'état civil ; France métropolitaine.

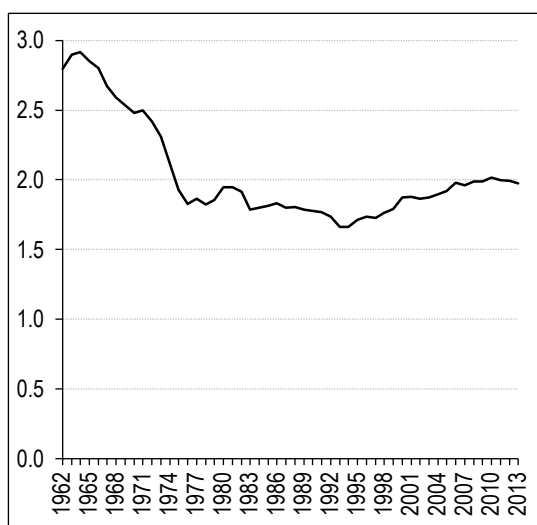
2. L'indice conjoncturel de fécondité (ICF) se calcule comme la somme des taux de fécondité par âge. Il correspond au nombre moyen d'enfants qu'aurait une femme au cours de sa vie si la probabilité de donner naissance à un âge donné correspond au taux de fécondité de cet âge.

Figure III – Évolution des logarithmes des taux de mortalité à différents âges de 1962 à 2013



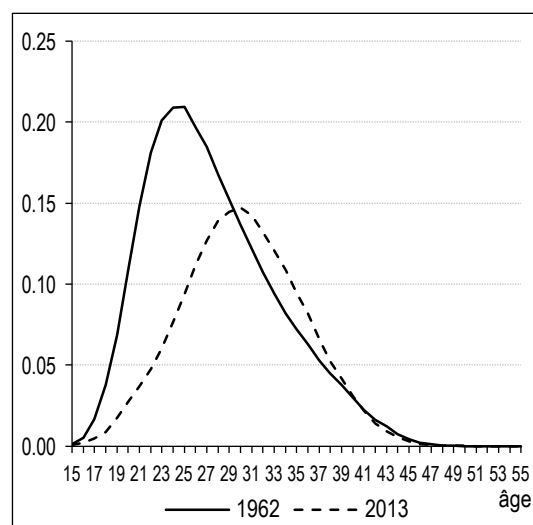
Source et champ : Insee, estimations de population et statistiques de l'état civil ; France métropolitaine.

Figure IV – Évolution de l'indice conjoncturel de fécondité de 1962 à 2013



Source et champ : Insee, estimations de population et statistiques de l'état civil ; France métropolitaine.

Figure V – Taux de fécondité par âge en 1962 et en 2013

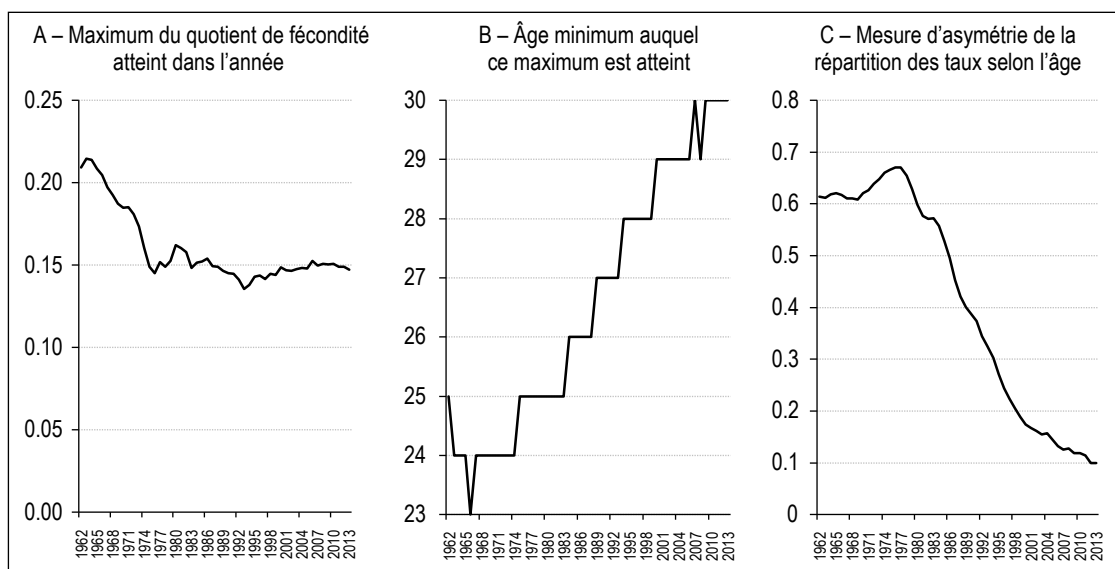


Source et champ : Insee, estimations de population et statistiques de l'état civil ; France métropolitaine.

Le taux de fécondité à un âge donné est défini comme le rapport entre le nombre d'enfants nés des femmes de cet âge et le nombre de femmes du même âge sur l'année considérée. Ce nombre de femmes correspond au nombre de femmes au 1^{er} janvier de l'année auquel on ajoute la moitié correspondante du solde migratoire et dont on soustrait la moitié correspondante des décès enregistrés dans l'année. Le profil des taux de fécondité par âge suit une courbe en cloche : la probabilité d'avoir un enfant dans l'année augmente avec l'âge à partir de 15 ans jusqu'à atteindre un pic, puis diminue continûment jusqu'à devenir nulle ou presque aux alentours de 50 ans.

Avec le temps, cette distribution par âge a tendance à se décaler vers la droite : l'âge auquel le maximum de fécondité est atteint augmente (figure V). En 1970, le taux de fécondité était maximal à 24 ans alors qu'en 2013, ce maximum est atteint à l'âge de 30 ans. Le niveau du maximum de fécondité atteint dans l'année n'évolue quant à lui guère depuis le milieu des années 1970 : il fluctue autour de 0.15. Le pic de fécondité se déplaçant vers la droite, la distribution des taux par âge devient de plus en plus symétrique, comme en témoigne la mesure d'asymétrie qui diminue rapidement vers 0 (figure VI).

Figure VI – Évolution du maximum de fécondité, de l'âge auquel le maximum de fécondité est atteint et de l'asymétrie de la distribution des taux de fécondité par âge, de 1962 à 2013

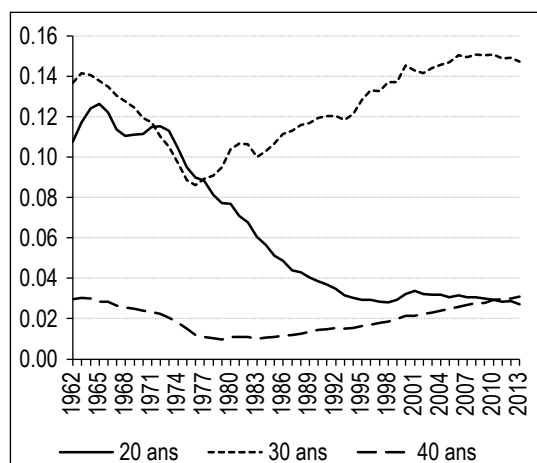


Source et champ : Insee, estimations de population et statistiques de l'état civil ; France métropolitaine.

Contrairement aux taux de mortalité, les taux de fécondité n'évoluent pas de façon régulière avec le temps. Par exemple, le taux de fécondité à 30 ans a diminué entre le début des années 1960 et le milieu des années 1970, et il augmente depuis lors avec un ralentissement à partir des années 2000. Le taux de fécondité à 20 ans diminue depuis les années 1970, mais à la fin des années 1990 il a connu un léger regain pendant quelques années, avant de diminuer à nouveau à un rythme beaucoup plus lent que lors des décennies précédentes. L'évolution n'est ni monotone ni linéaire, ce qui indique la difficulté à prolonger ces courbes dans le futur.

Pour résumer, le solde migratoire de France métropolitaine apparaît stable sur longue période, avec d'importantes fluctuations qui semblent difficiles à prévoir. La mortalité évolue depuis plusieurs décennies dans le même sens, avec une diminution quasi linéaire du logarithme des taux de mortalité à tous les âges et une diminution de l'écart d'espérance de vie entre les femmes et les hommes. L'évolution récente de la fécondité est plus complexe à cerner, mais il se dégage que l'ICF est stabilisé à un niveau moyen légèrement inférieur à 2 enfants par femme et que la distribution de la fécondité par âge se modifie continûment avec un déplacement du pic de fécondité vers des âges plus élevés et une distribution de plus en plus symétrique (figure VII). Dans la partie suivante, on propose une modélisation pour chacune des trois composantes de l'évolution de la population prenant en compte ces observations et s'inspirant de modèles déjà développés à l'international et qui y sont succinctement décrits.

Figure VII – Évolution des taux de fécondité à différents âges, de 1962 à 2013



Source et champ : Insee, estimations de population et statistiques de l'état civil ; France métropolitaine.

3. Méthodes et modélisations

Dans la suite de l'article, on utilisera les notations suivantes :

$P(a,n,s)$: nombre de personnes au 1^{er} janvier de l'année n , de sexe s nées l'année $n-a$;

$D(a,n,s)$: nombre de décès durant l'année n , de personnes de sexe s nées l'année $n-a$;

$N(a,n,s)$: nombre d'enfants de sexe s nés vivants durant l'année n et dont la mère est née l'année $n-a$;

$M(a,n,s)$: nombre d'entrées en France métropolitaine moins nombre de sorties, durant l'année n , de personnes de sexe s et nées l'année $n-a$. Cela correspond au solde migratoire de l'année n , pour les personnes de sexe s nées l'année $n-a$.

Pour faciliter les notations ensuite, on définit $P(0,n,s)$ comme le nombre de naissances vivantes d'enfants de sexe s l'année n . $D(0,n,s)$ et $M(0,n,s)$ correspondent respectivement, pour chaque année n et sexe s , au nombre de décès de bébés nés durant l'année n et au nombre de nouveau-nés entrés moins le nombre de nouveau-nés sortis du territoire. On supposera que les âges des femmes à l'accouchement se situent entre 15 et 55 ans inclus, et de ce fait on considérera que $N(a,n,s)=0$ pour $a \leq 14$ et $a \geq 56$.

On définit de plus les populations à risque, pour les décès et les naissances. Les populations à risque sont comptées en personnes-années et dépendent du nombre de personnes observées mais également de la période de temps sur laquelle ces personnes sont présentes. Pour les décès, cela correspond à la population au 1^{er} janvier de l'année à laquelle on ajoute la moitié du solde migratoire (si l'on considère que les flux d'entrées et de sorties sont répartis uniformément tout au long de l'année) :

$$R_D(a,n,s) = P(a,n,s) + 0.5M(a,n,s), \text{ si } a \geq 1$$

$$R_D(0,n,s) = 0.5P(0,n,s) + 0.5M(0,n,s),$$

pour $a = 0$.

Pour les naissances, le nombre de personnes-années à risque est le nombre moyen de femmes sur l'année, en supposant le flux des migrations et des décès uniformes :

$$R_N(a,n) = P(a,n,\text{femmes}) +$$

$$0.5M(a,n,\text{femmes}) - 0.5D(a,n,\text{femmes}).$$

On note de plus $M(n) = \prod_{a,s} M(a,n,s)$,

$$N(a,n) = N(a,n,\text{filles}) + N(a,n,\text{garçons})$$

$$\text{et } N(a,n) = \prod_a N(a,n,s).$$

Lorsqu'on notera des lois normales, on indiquera la moyenne et l'écart type (et non la variance).

3.1. Migrations

On projette directement le solde migratoire total à l'aide d'un modèle autorégressif d'ordre 1, où M_{lt} représente le solde migratoire de long terme et ε_M un bruit blanc :

$$M(n) = M_{lt} + \rho_M (M(n-1) - M_{lt}) + \varepsilon_M(n)$$

$$\varepsilon_M(n) \stackrel{i.i.d.}{\sim} N(0, \sigma_M)$$

Afin d'avoir un processus stationnaire, on impose la contrainte $|\rho_M| \leq 1$. Cette modélisation reflète le fait que l'on estime que le solde migratoire va continuer à être stable en moyenne et va osciller autour d'une tendance de long terme. L'amplitude des oscillations possibles pour l'avenir est déterminée par les amplitudes passées. De plus, on fixe un *a priori* très informatif sur la tendance de long terme en supposant, comme c'est le cas dans les travaux de Blanpain & Buisson (2016a), que ce niveau peut être estimé à partir du solde migratoire moyen sur la période récente, à savoir 80 000 personnes. La loi *a priori* pour la tendance de long terme est alors $M_{lt} \sim N(80\,000, 10\,000)$. Les paramètres M_{lt} , ε_M , ρ_M et σ_M sont estimés par inférence bayésienne à partir des données du solde migratoire sur la période 1995-2013.

Pour projeter le solde migratoire total, on tire aléatoirement 1 000 fois les paramètres du modèle selon leur distribution *a posteriori* et pour chaque jeu de paramètres on simule l'évolution du solde migratoire selon le processus autorégressif d'ordre 1. Une fois projeté le solde migratoire, on décompose ce dernier par sexe et âge selon des taux fixes calculés à partir de la répartition du solde migratoire par sexe et âge sur la période récente, et lissés, comme dans Blanpain & Buisson (2016a).

3.2. Mortalité

Comme on l'a vu, les logarithmes des taux de mortalité par âge semblent évoluer linéairement avec le temps. Néanmoins, l'évolution des taux de mortalité dans le temps se fait à un rythme différent pour chaque âge. On modélise directement le nombre de décès observés par une loi de Poisson, qui dépend du taux de mortalité et de la population à risque. Cette dernière correspond au nombre de personnes-années présentes en France métropolitaine l'année considérée. La loi de Poisson est couramment utilisée pour modéliser un nombre d'événements arrivant sur une période de temps donné. Elle est souvent retenue pour modéliser le nombre de décès dans les travaux de démographie. On retient la modélisation suivante, développée par Bryant & Zhang (2014), où $\mu_D(a, n, s)$ correspond au

taux de mortalité de l'année n , des personnes de sexe s et d'âge a :

$$D(a, n, s) \sim \text{Poisson}(\mu_D(a, n, s) R_D(a, n, s))$$

$$\log(\mu_D(a, n, s)) = \beta^0 + \beta_a^{\text{âge}} + \beta_{a,s}^{\text{âge:sexe}} + \beta_{a,n}^{\text{âge:année}} + \varepsilon_{D,1}(a, n, s)$$

Les $\varepsilon_{D,1}$ sont des termes d'erreur indépendants et identiquement distribués selon une loi normale centrée et d'écart-type $\sigma_{D,1}$. Le paramètre β^0 est une constante, le paramètre $\beta_a^{\text{âge}}$ donne la répartition moyenne selon l'âge du logarithme des taux de mortalité. Enfin, il y a deux termes croisant deux dimensions : $\beta_{a,s}^{\text{âge:sexe}}$ qui permet d'estimer un effet du sexe spécifique à chaque âge et un paramètre $\beta_{a,n}^{\text{âge:année}}$ qui est un effet du temps spécifique à chaque âge. On notera donc que l'évolution temporelle du logarithme des taux de mortalité par âge est la même pour les femmes et les hommes puisqu'on n'a pas spécifié de terme croisant l'année et le sexe. On a ce faisant voulu limiter le nombre de paramètres à estimer. L'introduction d'un terme supplémentaire croisant année et sexe conduit à une mauvaise estimation des distributions *a posteriori*, en raison d'une non-convergence des chaînes de Markov. À un troisième niveau, on modélise certains des paramètres par des modèles linéaires dynamiques. On peut ainsi décomposer l'évolution temporelle du paramètre $\beta_{a,n}^{\text{âge:année}}$, par âge, en un niveau ($\theta_{a,n}^{\text{âge:année}}$) et une tendance ($\delta_{a,n}^{\text{âge:année}}$) :

$$\beta_{a,n}^{\text{âge:année}} = \theta_{a,n}^{\text{âge:année}} + \eta(a, n)$$

$$\theta_{a,n}^{\text{âge:année}} = \theta_{a,n-1}^{\text{âge:année}} + \delta_{a,n}^{\text{âge:année}} + \nu(a, n)$$

$$\delta_{a,n}^{\text{âge:année}} = \delta_{a,n-1}^{\text{âge:année}} + \omega(a, n)$$

Les termes η , ν et ω sont des termes d'erreurs indépendants, suivant une loi normale centrée.

Pour projeter dans le futur les taux de mortalité par âge, il suffit donc, une fois estimée la distribution *a posteriori* de l'ensemble des paramètres du modèle, de générer des nouveaux termes de tendance, puis des nouveaux termes de niveaux et enfin des nouveaux paramètres $\beta_{a,n}^{\text{âge:année}}$, jusqu'à l'horizon souhaité.

3.3. Fécondité

Pour la fécondité, on procède en trois étapes. On projette en premier lieu l'ICF selon un modèle autorégressif d'ordre 1. L'ONU utilise la même méthode pour sa troisième phase d'évolution de la fécondité, en supposant que l'ICF tend vers 2.1 pour tous les pays (Alkema *et al.*, 2010). Par rapport à la méthode de l'ONU, on choisit de plus d'estimer les paramètres du modèle par

inférence bayésienne, et non par maximum de vraisemblance. On reste ainsi dans un cadre entièrement bayésien, pour l'ensemble de nos estimations et de nos projections. Le modèle est le suivant :

$$ICF(n) = ICF_t + \rho_F (ICF(n-1) - ICF_t) + \varepsilon_F(n)$$

où

$$ICF(n) = \prod_{a=15}^{55} \frac{N(a, n, filles) + N(a, n, garçons)}{R_F(a, n)}$$

est l'indicateur conjoncturel de fécondité de l'année n . Comme pour le solde migratoire, on simule, après estimation par inférence bayésienne, 1 000 trajectoires possibles d'évolution de cet indicateur jusqu'à l'horizon souhaité.

La deuxième étape consiste à projeter, indépendamment de la projection de l'ICF, les taux de fécondité par âge μ_F . Ceux-ci sont définis, comme dans le cas de la mortalité, *via* la modélisation des naissances par un processus de Poisson :

$$N(a, n) \sim \text{Poisson}(\mu_F(a, n) R_F(a, n))$$

où l'on rappelle que $N(a, n)$ correspond au nombre de naissances l'année n données par des mères nées l'année $n-a$. Suivant la méthode proposée par Bijak *et al.* (2015) inspirée de la méthode de Lee-Carter, on modélise ensuite le logarithme du taux de fécondité comme la somme d'un effet fixe de l'âge, d'un effet du temps dont l'intensité et la direction sont différentes pour chaque âge et d'un effet de génération :

$$\log(\mu_F(a, n)) = \alpha_a + \beta_a \kappa_n + \gamma_{n-a} + \varepsilon_{F,1}(a, n)$$

L'effet temporel κ et l'effet de génération γ évoluent selon des processus autorégressifs d'ordre 1 :

$$\kappa_n = \varphi_0 + \varphi_1 \kappa_{n-1} + \xi(n)$$

$$\gamma_{n-a} = \Psi_0 + \Psi_1 \gamma_{n-a-1} + \zeta(n)$$

où les termes d'erreurs ξ et ζ suivent des lois normales d'espérance nulle. Cette fois encore, tous les paramètres sont estimés par inférence bayésienne, pour ensuite produire 1 000 simulations des taux de fécondité, pour chaque âge et chaque année future. Ces taux projetés prolongent des tendances linéaires, même si les paramètres φ_1 et ψ_1 peuvent, s'ils sont de norme strictement plus petite que 1, amener à annuler l'effet temporel ou l'effet de génération à long terme. Les estimations donnent une distribution *a posteriori* de φ_1 et ψ_1 qui sont très proches de 1. Il en résulte qu'à moyen terme, les taux de fécondité deviennent anormalement élevés pour certains âges, ce qui conduit à des ICF bien plus élevés que ceux projetés dans la première étape.

La troisième étape consiste alors à corriger les taux de fécondité par âge. Ils sont calés, chaque année, sur l'ICF projeté. Pour cela on multiplie simplement l'ensemble des taux, une année donnée, par une constante. Notons qu'aucune contrainte n'a été ajoutée sur l'âge moyen à la maternité, alors que les projections de l'Insee retiennent un plafond à 32 ans issu des consultations d'experts (voir Blanpain, ce numéro).

La descendance finale est basée sur les taux de fécondité d'une génération donnée de femmes. Comme l'ICF, l'indicateur est une synthèse de taux de fécondité à différents âges. En revanche, contrairement à l'ICF qui est un indicateur transversal, la descendance finale est un indicateur longitudinal qui nécessite d'observer la vie féconde d'une génération entière avant de le calculer. Or nous ne disposons que d'un nombre limité de points d'observations dans le passé. C'est pourquoi nous avons décidé, à l'instar de nombreux auteurs, de modéliser et de projeter l'ICF. Notons que l'espérance de vie est également un indicateur transversal.

3.4. Projections par la méthode des composantes

La méthode des composantes permet de faire évoluer la population d'une année sur l'autre en remarquant que la population au 1^{er} janvier d'une année donnée est égale à la population au 1^{er} janvier de l'année précédente, à laquelle on ajoute le nombre de naissances moins le nombre de décès ayant eu lieu l'année précédente, et le solde migratoire, ce que retracent les équations suivantes :

$$P(a, n, s) = P(a-1, n-1, s) - D(a-1, n-1, s) + M(a-1, n-1, s)$$

pour $a \geq 1$ et $P(0, n, s) = N(n, s)$.

Les nombres de décès et de naissances sont obtenus chaque année par tirage aléatoire selon une loi de Poisson (voir les modélisations). Pour cela, il faut déterminer les personnes à risque pour les décès et les femmes à risque pour les naissances. On commence par calculer les décès pour chaque âge, excepté pour les décès des nouveau-nés. On en déduit ensuite les femmes à risque à chaque âge entre 15 ans et 55 ans (on doit en effet pour cela connaître le solde migratoire et le nombre de décès). Enfin, on calcule le nombre de décès des nouveau-nés. La répartition du nombre de naissances une année donnée, entre naissances de garçons et naissances de filles est déterminée par le sex-ratio que l'on fixe à 1.05, conformément aux constats passés.

3.5. Validations des modèles

Une manière de tester les modèles utilisés est de séparer les données portant sur le passé en deux catégories : une partie, environ deux tiers, est utilisée pour estimer les modèles et la partie restante, environ un tiers, est utilisée pour confronter les estimations du modèle à la réalité.

Pour la mortalité, on choisit d'estimer le modèle sur la période 1962-1995 et de comparer les résultats sur la période 1996-2013. Pour la fécondité on estime les modèles sur la période 1975-2000 et on compare les résultats sur la période 2001-2013. On constate que le logarithme des taux de mortalité est correctement projeté aux grands âges (à partir de 35-40 ans environ), mais que la modélisation retenue donne des diminutions moins rapides de ces taux que ce qui est observé en réalité. Cela est dû au fait que pour les très jeunes âges, le logarithme du taux de mortalité est légèrement concave et non linéaire. De plus, pour les jeunes adultes, les taux de mortalité ont plus ou moins stagné dans les années 1980 et 1990 avant de baisser fortement. Le modèle n'a pas pu anticiper cette baisse soudaine.

En ce qui concerne la fécondité, l'intervalle de confiance à 95 % des projections probabilistes de l'ICF contient bien l'ICF observé. Mais lorsque l'on regarde la distribution des taux de fécondité par âge, on se rend compte que la méthode utilisée conduit à une distribution plus resserrée que ce qui est réellement observé (figure VIII). La déformation de la distribution des taux de fécondité par âge est donc un peu trop forte dans nos projections.

4. Résultats des projections probabilistes bayésiennes pour la France jusqu'en 2070

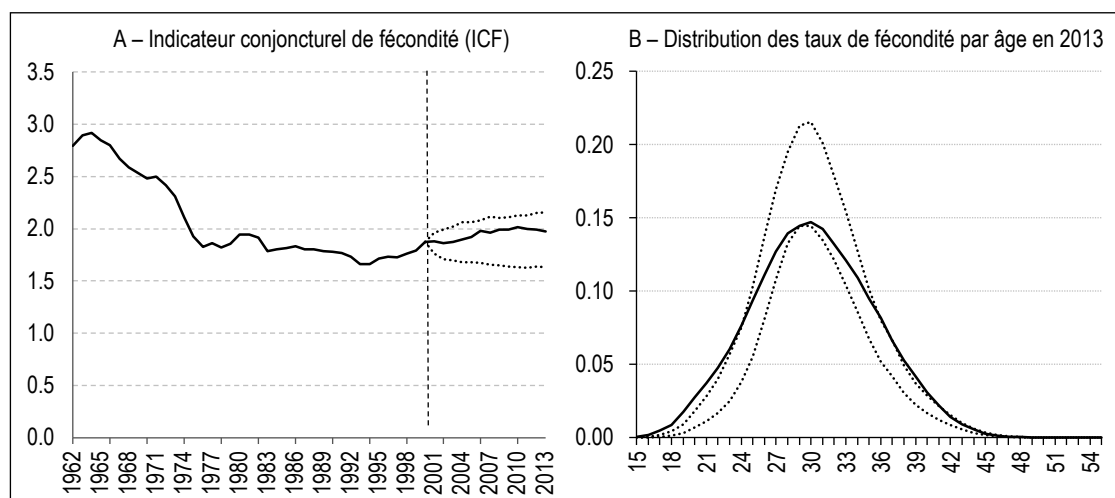
Les paramètres des modèles pour le solde migratoire, la mortalité et la fécondité ont été estimés par inférence bayésienne à l'aide du logiciel libre *Stan* et du package R *demest* publié par l'Institut statistique de Nouvelle Zélande³. Pour chacun des paramètres on a simulé 1 000 valeurs selon leur loi *a posteriori*. On a ensuite généré 1 000 trajectoires d'évolution possible pour le solde migratoire, les taux de mortalité par sexe et âge et les taux de fécondité par âge. Au final on peut obtenir 1 000 estimations de n'importe quel indicateur démographique dérivé de ces trois composantes, dont notamment la taille de la population totale. On en déduit alors des intervalles de confiance à 95 % ou à 80 % qui contiennent respectivement 95 % ou 80 % des estimations.

4.1. Projections des migrations : une incertitude forte et constante

Le solde migratoire projeté suit une évolution stable car cela a été spécifié ainsi dans le modèle. La médiane des 1 000 trajectoires possibles diminue dans les premières années de projections puis se stabilise rapidement à 79 000 (figure IX). L'intervalle de confiance lui aussi reste constant au cours du temps : avec une probabilité de 95 % le solde migratoire se maintiendra entre 29 000 et 129 000 chaque année. Cette amplitude est due aux larges fluctuations observées dans le passé, elle dépasse

3. <https://github.com/StatisticsNZ/demest>

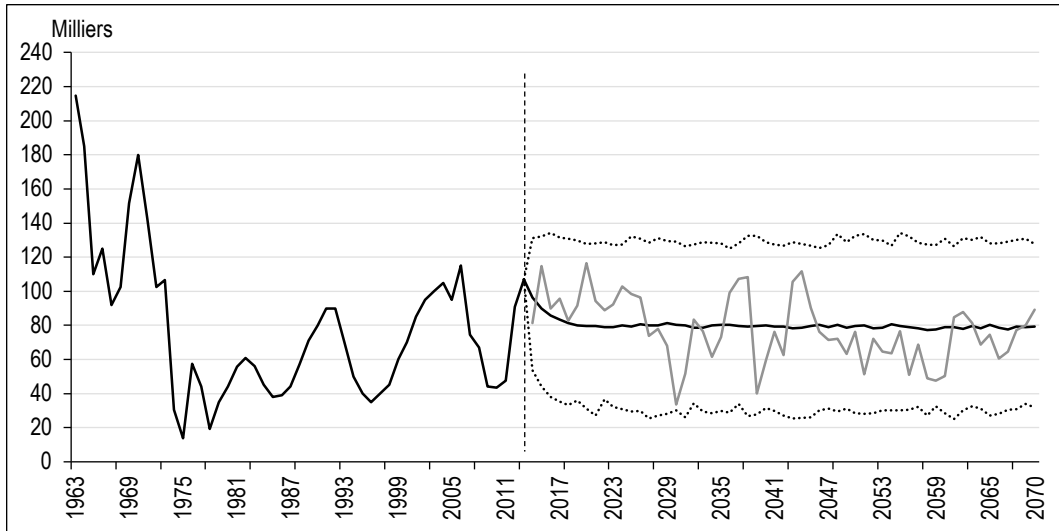
Figure VIII – ICF et taux de fécondité par âge, observés (1962-2013) et projetés (2001-2013)



Note : en pointillés sont indiqués les quantiles d'ordre 2.5 % et 97.5 % des projections probabilistes et en trait plein l'ICF et les taux de fécondité réels (de 1962 à 2013).

Source et champ : Insee, estimations de population et statistiques de l'état civil (taux de fécondité), calculs de l'auteur (projections probabilistes) ; France métropolitaine.

Figure IX – Solde migratoire passé et projeté



Note : en pointillés sont indiqués les quantiles d'ordre 2.5 % et 97.5 % et en trait plein la médiane des distributions *a posteriori*. La courbe en gris clair représente une des 1 000 simulations.
Source et champ : Insee, estimations de population et statistiques de l'état civil (1962-2013), calculs de l'auteur (2013-2070) ; France métropolitaine.

légèrement les minimum et maximum observés respectivement en 1996 et 2006.

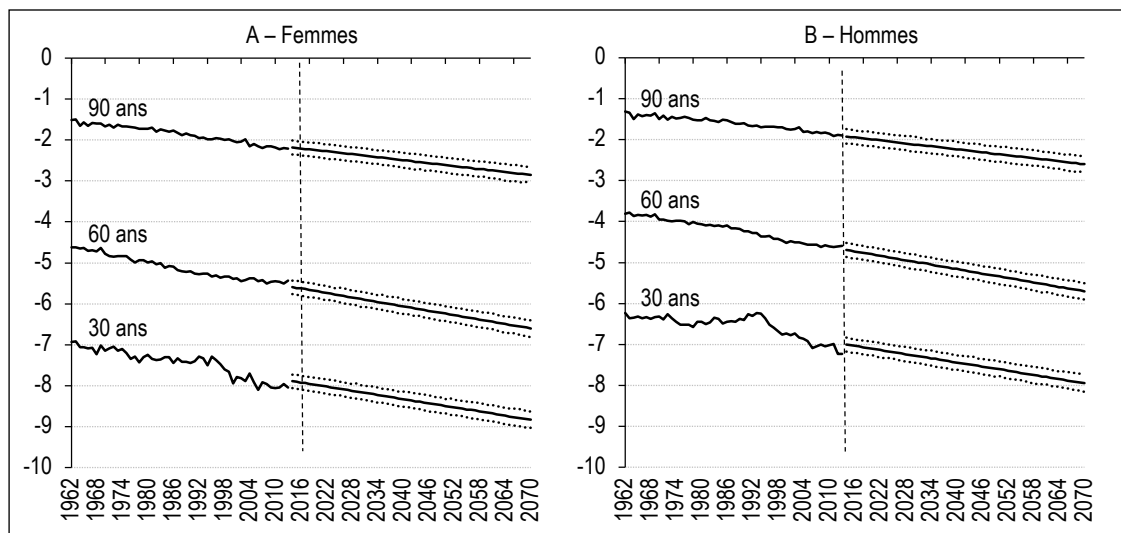
4.2. Projections de la mortalité : peu d'incertitude compte tenu des évolutions passées

Le modèle pour la mortalité prévoit que les taux de mortalité par âge continueront à diminuer linéairement, selon la même tendance pour les hommes et pour les femmes (figure X). L'incertitude sur les taux de mortalité projetés n'augmente presque pas avec le temps. Cela provient de ce que les variances des erreurs de niveau et de tendance ν et ω sont très faibles comparées à

la variance du terme d'erreur η . Les erreurs ne s'accroissent donc pas avec le temps. Cela est dû au fait que les tendances observées sont très linéaires.

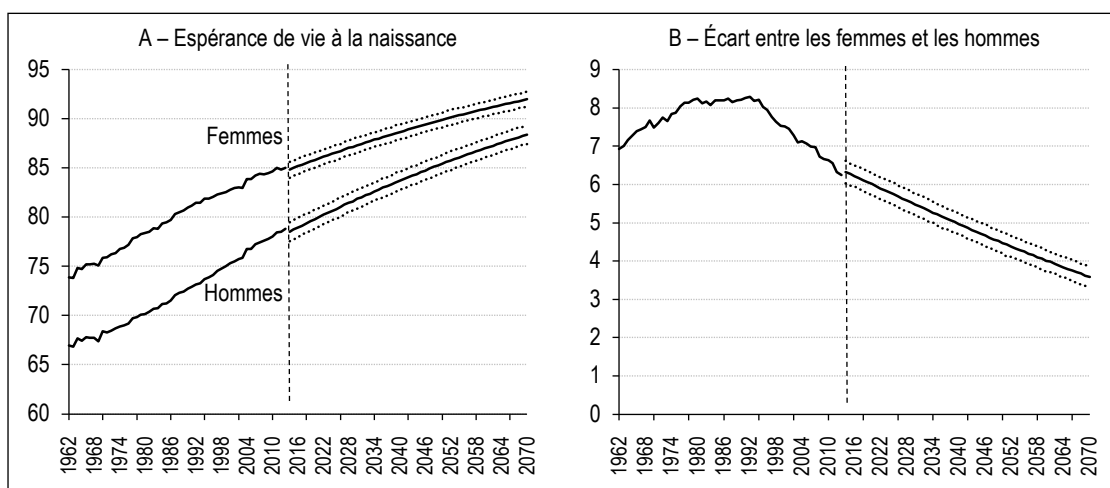
Du fait de la diminution constante des taux de mortalité, l'espérance de vie continuera à croître dans les décennies qui viennent, pour les hommes comme pour les femmes. Les résultats du modèle indiquent que l'espérance de vie à la naissance des femmes sera, avec une probabilité de 95 %, comprise entre 91.2 ans et 92.8 ans en 2070 et celle des hommes entre 87.4 ans et 89.4 ans (figure XI). L'écart d'espérance de vie entre les femmes et les hommes va probablement continuer

Figure X – Évolution du logarithme des taux de mortalité par âge, estimés et projetés



Note : en pointillés sont indiqués les quantiles d'ordre 2.5 % et 97.5 % et en trait plein la médiane des distributions *a posteriori*.
Source et champ : Insee, calculs de l'auteur, France métropolitaine.

Figure XI – Évolution de l'espérance de vie estimée et projetée et de l'écart d'espérance de vie entre les femmes et les hommes



Note : en pointillés sont indiqués les quantiles d'ordre 2.5 % et 97.5 % et en trait plein la médiane des distributions *a posteriori*.
 Source et champ : Insee, estimations de population et statistiques de l'état civil (1962-2013), calculs de l'auteur (2013-2070) ; France métropolitaine.

à se résorber pour atteindre 3.6 ans en 2070 (entre 3.3 ans et 3.9 ans avec une probabilité de 95 %).

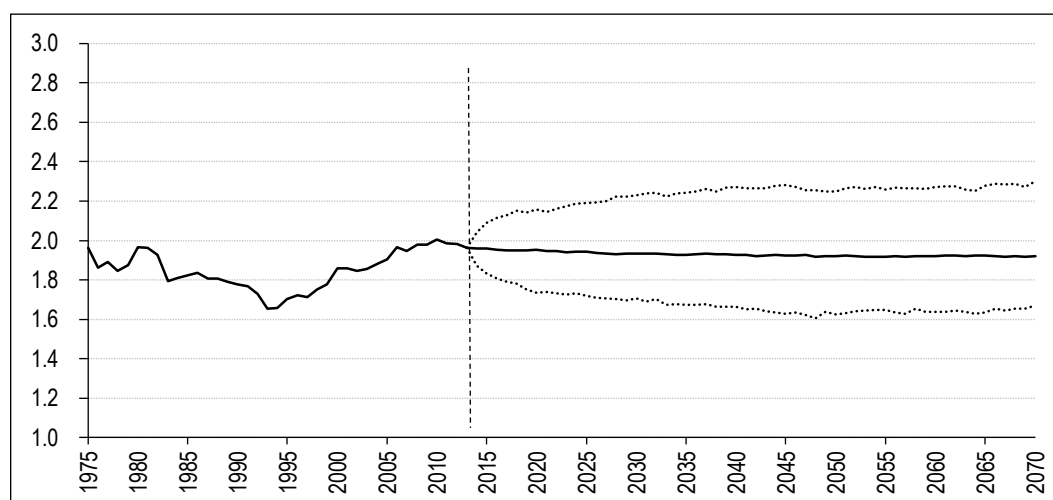
4.3. Projections de la fécondité : des maternités plus tardives et réparties de façon plus symétrique autour de l'âge modal

La médiane de l'ICF de long terme est de 1.93, légèrement en-dessous de la moyenne de la loi *a priori* fixée à 1.95 (figure XII). Selon la modélisation retenue, l'ICF sera compris avec une probabilité de 95 % entre 1.63 et 2.26 enfants par femme en 2070. Contrairement aux projections du solde migratoire et des taux de mortalité, l'intervalle de confiance à 95 % devient de plus en plus large avec le temps. L'incertitude sur la

fécondité future devient donc de plus en plus grande malgré le fait d'avoir fixé un ICF de long terme dans la modélisation.

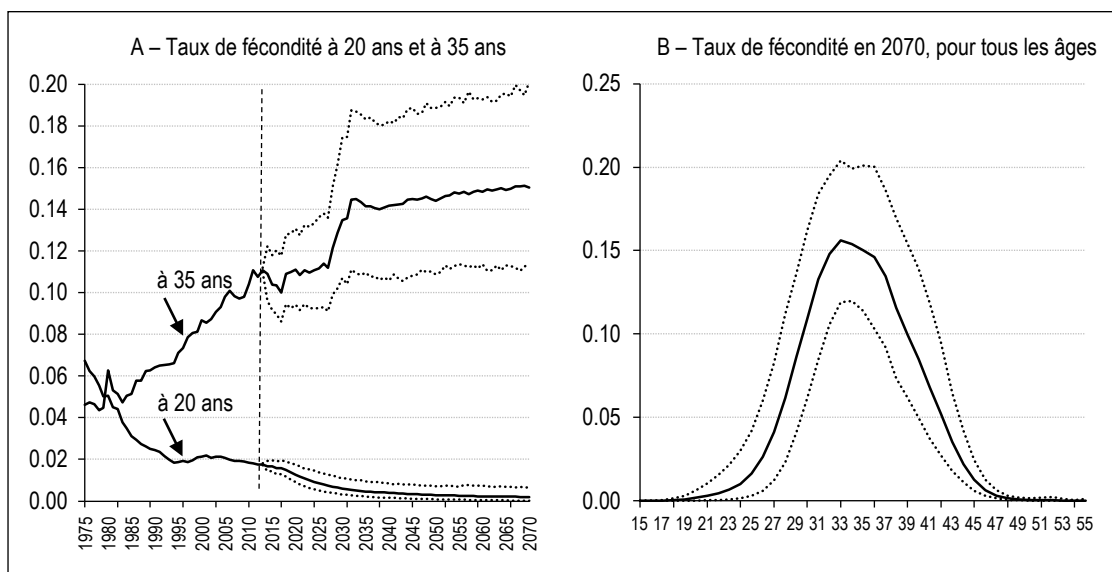
Les taux de fécondité par âge commencent à se stabiliser à partir de 2050 (figure XIII). L'âge moyen à la maternité augmente plutôt rapidement jusqu'en 2040 environ, puis l'augmentation se poursuit mais à un rythme de moins en moins rapide, pour atteindre une valeur comprise entre 32.2 et 35.9 ans en 2070 (intervalle de confiance à 95 %). La distribution des taux de fécondité par âge se décale alors de plus en plus vers la droite et devient de plus en plus symétrique, comme le montre l'évolution de l'indicateur d'asymétrie dont la médiane tend vers 0 (figure XIV).

Figure XII – Évolution de l'indice conjoncturel de fécondité, estimé et projeté



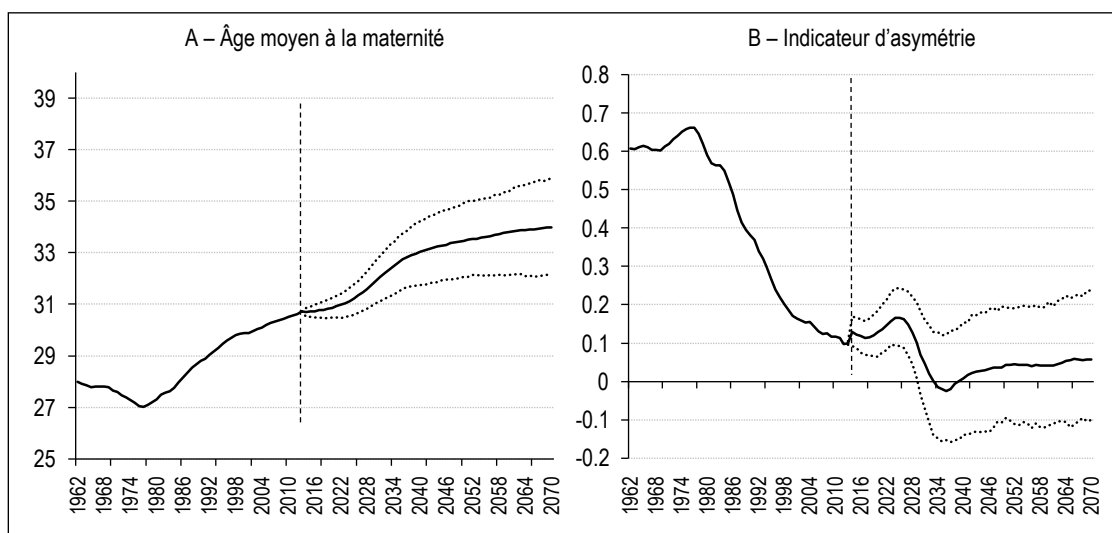
Note : en pointillés sont indiqués les quantile d'ordre 2.5 % et 97.5 % et en trait plein la médiane des distributions *a posteriori*.
 Source et champ : Insee, estimations de population et statistiques de l'état civil (1962-2013), calculs de l'auteur (2013-2070) ; France métropolitaine.

Figure XIII – Évolution des taux de fécondité estimés et projetés



Note : en pointillés sont indiqués les quantiles d'ordre 2.5 % et 97.5 % et en trait plein la médiane des distributions *a posteriori*.
Source : Insee, calculs de l'auteur ; France métropolitaine.

Figure XIV – Évolution de l'âge moyen à la maternité et de l'asymétrie de la distribution des taux de fécondité par âge



Note : en pointillés sont indiqués les quantile d'ordre 2.5 % et 97.5 % et en trait plein la médiane des distributions *a posteriori*.
Source et champ : Insee, estimations de population et statistiques de l'état civil (1962-2013), calculs de l'auteur (2013-2070) ; France métropolitaine.

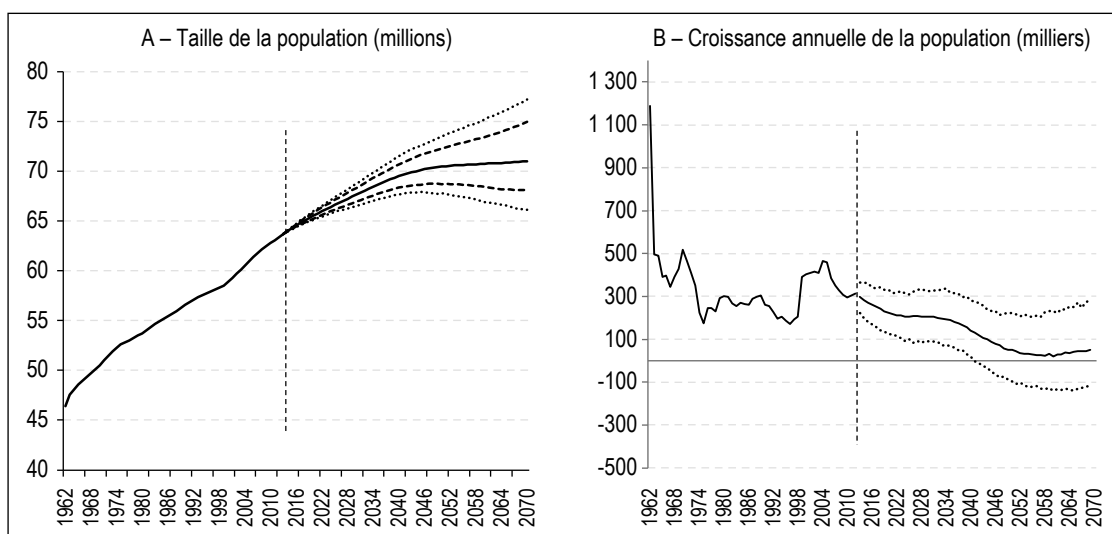
4.4. Projections de la population totale : une croissance vraisemblablement forte d'ici à 2040, et beaucoup plus faible ensuite

La population totale de la France métropolitaine va continuer à augmenter pour atteindre en 2070 un niveau compris entre 66.1 millions et 77.2 millions avec une probabilité de 95 % et entre 68.1 millions et 75.0 millions avec une probabilité de 80 % (figure XV). La projection médiane correspond à un niveau de 71.0 millions d'habitants en 2070. La population de France métropolitaine pourrait donc augmenter continûment tout au long des

cinquante prochaines années ou bien augmenter et commencer à décliner vers 2050. Il y a, selon la modélisation retenue ici, une probabilité de 1 % pour que la population commence à baisser dès 2040 (c'est-à-dire que la population atteigne son maximum en 2040) et une probabilité de 19 % pour 2050. L'incertitude sur la taille de la population d'après la modélisation retenue est assez faible jusque vers 2040-2050, puis augmente plus fortement avec les années suivantes.

La structure de la population va également être modifiée, comme en témoigne la pyramide des

Figure XV – Évolution passée et projetée de la taille totale de la population et de la croissance annuelle de la population



Note : en pointillés sont indiqués les quantiles d'ordre 2.5 % et 97.5 %, en tirets ceux d'ordre 10 % et 90 % et en trait plein la médiane des distributions *a posteriori*.

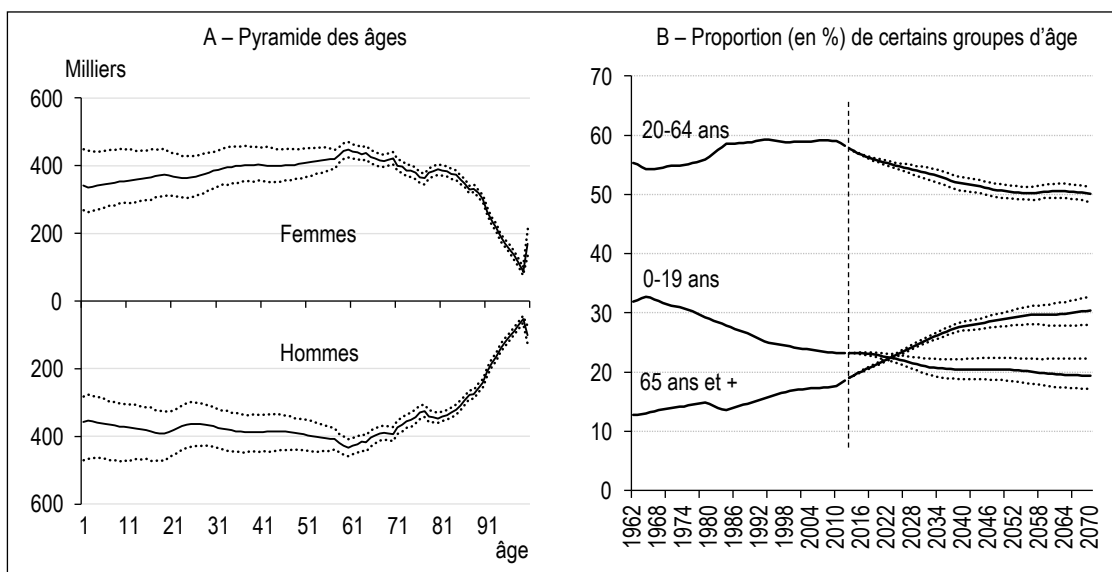
Source et champ : Insee, estimations de population et statistiques de l'état civil (1962-2013), calculs de l'auteur (2013-2070) ; France métropolitaine.

âges en 2070 dont la base est bien plus droite et fine que la pyramide des âges actuelle. Certains groupes d'âges vont ainsi baisser en proportion, notamment les plus jeunes (figure XVI) : la part des 0-19 ans va continuer à diminuer lentement pour atteindre en 2070 un niveau médian de 19 %, celles des 20-64 ans suit le même profil, avec en 2070 un niveau médian de 50 %. Au contraire, la part des 65 ans et plus dans la population va probablement continuer d'augmenter pour être en 2070 plus importante que la part des moins de 20 ans. Elle est passée de 13 % en 1962 à 19 % en 2013 et pourrait atteindre, avec

une probabilité de 95 %, entre 28 % et 33 % de la population en 2070.

La population va donc continuer à vieillir. L'âge médian de la population, qui est en 2013 de 41 ans, pourrait être, avec une probabilité de 95 %, compris entre 44 et 50 ans en 2070. En conséquence, le ratio entre les personnes de 65 ans et plus et les personnes de 20 à 64 ans risque d'augmenter fortement dans les années à venir. L'augmentation rapide et linéaire de ce ratio entre aujourd'hui et le début des années 2040 est principalement due au vieillissement

Figure XVI – Pyramide des âges en 2070 et évolution de la proportion de certains groupes d'âge



Note : en pointillés sont indiqués les quantile d'ordre 2.5 % et 97.5 % et en trait plein la médiane des distributions *a posteriori*.

Source et champ : Insee, estimations de population et statistiques de l'état civil (1962-2013), calculs de l'auteur (2013-2070) ; France métropolitaine.

des générations nombreuses nées pendant le baby-boom. En effet, les personnes nées en 1946, début du baby-boom, ont eu 65 ans en 2011 et celles nées à la fin du baby-boom en 1975 auront 65 ans en 2040. Le ratio entre les 65 ans ou plus et les 20-64 ans, aujourd'hui de 0.33, atteindrait selon les modèles utilisés, une valeur comprise entre 0.56 et 0.67 avec une probabilité de 95 % en 2070 (figure XVII).

On peut comparer ces projections probabilistes aux projections déterministes réalisées par l'Insee. Les projections déterministes concernant la France métropolitaine ne concernent que la période 2013-2050⁴. Le scénario central retenu conduit à une population légèrement plus élevée que la médiane de nos projections probabilistes : en 2050 la population de France métropolitaine atteindrait 71.7 millions d'habitants selon la première projection et 70.5 millions selon la deuxième. Par ailleurs, l'intervalle de confiance estimé par les projections probabilistes est bien plus faible que l'intervalle entre le scénario de population haute et le scénario de population basse, qui sont les scénarios extrêmes des projections déterministes. L'écart entre les deux scénarios déterministes extrêmes est de 11.1 millions d'habitants en 2050, alors que l'intervalle de confiance des projections déterministes cette même année est de 5.7 millions pour l'intervalle de confiance à 95 % et de 3.6 millions pour celui à 80 %.

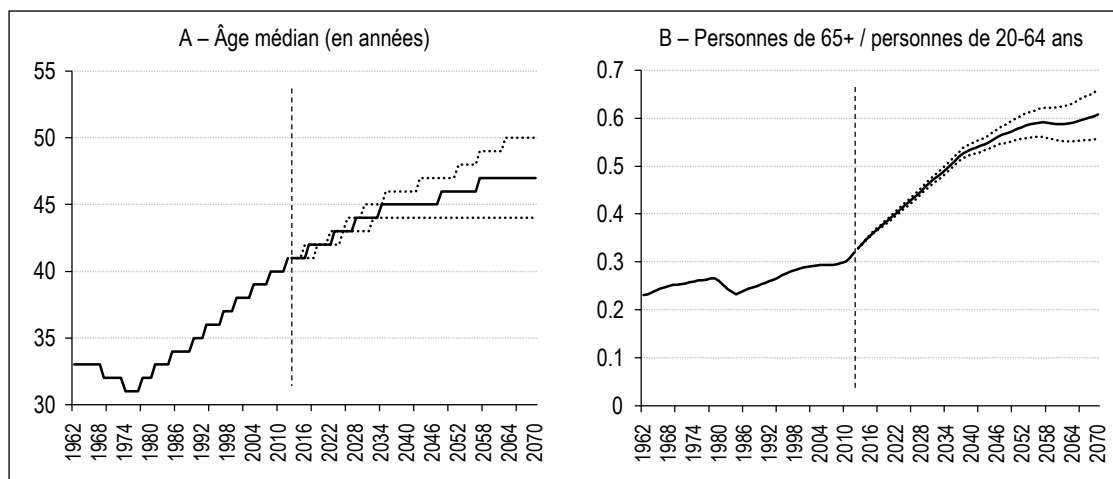
4.5. Discussion

D'après les modélisations décrites dans cet article et les simulations effectuées, la population de France métropolitaine devrait continuer

à augmenter dans les décennies à venir. Elle pourrait pourtant commencer à décroître avant 2070, avec une probabilité non négligeable, même si cette évolution est moins probable qu'une hausse ou une stabilisation. La structure de cette population sera également probablement modifiée : on s'attend à un vieillissement général de la population dû à l'augmentation de l'espérance de vie, à la stagnation de l'ICF et à l'arrivée continue des baby-boomers aux âges de la retraite. Le modèle utilisé pour projeter le solde migratoire est le plus simple des trois modélisations utilisées. L'absence de données détaillée par sexe et âge sur les entrées et les sorties empêchent d'utiliser une modélisation de Poisson faisant apparaître des taux, comme on l'a fait pour le nombre de décès et le nombre de naissances. De manière générale, les modèles de projection du solde migratoire sont moins sophistiqués et ont fait l'objet de moins d'effort de recherche que ceux concernant la mortalité et la fécondité, les données disponibles étant moins riches. On peut noter toutefois que certains pays disposant de registres, notamment la Nouvelle Zélande, ayant des données détaillées d'entrées et de sorties, commencent à proposer des modélisations avancées des phénomènes migratoires, prenant en compte de nombreux paramètres, comme par exemple le niveau d'éducation de la population (Bryant & Zhang, 2014). Notre modélisation étant assez simple, il en résulte que la plupart des évolutions passées du solde migratoire sont considérées comme du bruit. Ce bruit étant ensuite propagé dans le futur, les intervalles de confiance du solde migratoire projeté sont très larges et reflètent de ce fait notre

4. voir <https://www.insee.fr/fr/statistiques/2859843>

Figure XVII – Évolution de l'âge médian de la population et du ratio des personnes de 65 ans et plus sur les personnes de 20-64 ans



Note : en pointillés sont indiqués les quantiles d'ordre 2.5 % et 97.5 % et en trait plein la médiane des distributions *a posteriori*.
Source et champ : Insee, estimations de population et statistiques de l'état civil (1962-2013), calculs de l'auteur (2013-2070) ; France métropolitaine.

niveau d'incertitude sur l'évolution à venir des migrations. C'est pour cette raison que nous avons restreint l'estimation des paramètres du modèle (et donc en particulier de la variance du terme d'erreur) à la période 1995-2013, afin de ne pas prendre en compte les larges fluctuations du solde migratoire, trop anciennes. Estimer le modèle sur une plus longue période aurait conduit à une incertitude encore plus grande sur l'évolution future du solde migratoire.

Au contraire du solde migratoire, les évolutions de la mortalité sont très stables, et le modèle utilisé peut rendre compte de ces évolutions sans les considérer comme étant majoritairement du bruit. Il en résulte que les intervalles de confiance des taux de mortalité projetés et de l'espérance de vie projetée sont très étroits. Ceci peut être trompeur, car on pourrait croire que l'on est presque certain de ce qui va arriver. En réalité, il ne faut pas oublier que ces intervalles de confiance sur les niveaux futurs de la mortalité sont déterminés conditionnellement au fait que le modèle offre une bonne approche du réel. C'est bien en supposant que les tendances observées vont se poursuivre que l'on peut arriver à de tels niveaux de confiance concernant l'avenir de la mortalité. Malgré cela, le modèle utilisé ne prend pas en compte certaines spécificités de la mortalité en France. En premier lieu, il ne permet pas de projeter des évolutions du logarithme du quotient de mortalité à un âge donné différentes selon le sexe. De plus, il apparaît que les générations nées après la seconde guerre mondiale ont très peu de gain au niveau de la mortalité à âge donné par rapport aux générations précédentes, et ce quel que soit l'âge (Blanpain & Buisson, 2016a). La modélisation retenue ne permet pas de prendre en compte de tels effets de générations : les écarts à la tendance générale sont alors traités comme du bruit, mis dans les termes d'erreur, plutôt que comme un effet bien identifié. Les espérances de vie projetées auxquelles on a abouti sont donc un peu plus faibles que celles obtenues avec les projections de Blanpain & Buisson (2016a).

La modélisation de la fécondité est différente de celle du solde migratoire et de la mortalité. En effet, contrairement aux taux de mortalité, les taux de fécondité n'ont pas une évolution régulière au cours du temps. Ils peuvent augmenter puis diminuer, ou faire l'inverse et de ce fait se croiser. Prolonger les taux de fécondité selon des tendances linéaires amène ainsi à des situations qui apparaissent invraisemblables au vu d'autres indicateurs de la fécondité, comme l'ICF ou le maximum de fécondité atteint dans l'année, qui sont plus ou moins stables depuis 1975. L'idée

a alors été de prolonger dans un premier temps l'ICF, qui est un indicateur qui reflète le niveau de la fécondité, selon une méthode identique à celle utilisée pour projeter le solde migratoire. On a ensuite prolongé les taux de fécondité par âge selon la méthode de Wiśniowski *et al.* (2015), et on a modifié ces taux pour retrouver l'ICF projeté. On dispose ainsi d'une évolution assez réaliste des taux de fécondité par âge, dont la distribution se décale vers les âges plus élevés tout en devenant plus symétrique. Cette façon de procéder (projection d'un indicateur agrégé, puis ventilation par catégories détaillées) n'est pas nouvelle en soi et c'est aussi le schéma retenu par l'ONU. L'inconvénient est qu'il faut ici fixer un ICF de long terme et le niveau choisi joue bien évidemment sur les résultats.

* *
*

Les projections de population probabilistes offrent un nouvel éclairage sur l'évolution possible de la population. Elles permettent, sous certaines hypothèses de modélisation, de quantifier l'incertitude sur l'évolution à venir des indicateurs démographiques, et notamment sur l'évolution de la taille totale de la population. Elles présentent donc un avantage certain sur les projections déterministes basées sur des scénarios dont la probabilité d'occurrence n'est pas quantifiée. N'importe quel indicateur démographique, comme l'espérance de vie, l'âge moyen à la maternité ou la part des 65 ans et plus, peut être déterminé avec un certain degré de probabilité. Une des difficultés potentielles de l'interprétation des résultats provient du fait qu'il ne faut pas penser en termes ponctuels, mais plutôt en termes de distribution de probabilité, tout comme un dé, même pipé, ne peut se résumer à une seule de ces six faces. C'est plutôt en donnant la probabilité d'apparition de chaque face que l'on aura une bonne description du dé et de ce que l'on peut en attendre lorsqu'il sera lancé. Une fois cette difficulté surmontée, l'interprétation et l'utilisation de projections de population probabilistes offrent une grande liberté et une grande souplesse. *A contrario*, les résultats des projections déterministes deviennent compliqués à utiliser et leur communication difficile quand le nombre de scénarios envisagés est multiplié sous l'effet du croisement de plusieurs hypothèses.

Afin d'améliorer les méthodes utilisées dans cet article et par conséquent les résultats, plusieurs pistes sont envisageables. Il s'agit en premier

lieu de mieux comprendre les phénomènes migratoires et pour cela d'analyser les entrées de façon détaillée. Il serait aussi intéressant de se pencher sur des estimations des flux de sorties actuels et passés, ce qui est un travail assez récent en France, compte tenu des données disponibles. Pour la projection de la mortalité il serait utile d'intégrer un effet de génération et d'autoriser une évolution différente des taux de mortalité pour les femmes et les hommes. Plusieurs modèles sont pour cela envisageables mais s'il y a trop de paramètres, le risque de non identification du modèle ou de mauvaise convergence des chaînes de Markov servant à estimer les distributions *a posteriori* est important. Pour améliorer la projection des taux de fécondité par âge, on pourrait, comme cela a déjà été fait dans plusieurs travaux, trouver une modélisation paramétrique de la distribution des taux selon l'âge. Il suffirait alors, et ce n'est pas forcément une chose facile, de prolonger ces paramètres comme dans le cas de la modélisation de Lee-Carter, en détectant les régularités et les tendances dans l'évolution de ces paramètres. La distribution de la loi Bêta est une modélisation possible, mais sa forme arrondie ne représenterait pas bien les données. La loi Gamma à l'intérêt de mieux refléter la distribution des taux de fécondité, mais elle est définie sur un support ouvert à droite. Il faudrait alors la tronquer pour ne pas avoir de résultats irréalistes. La fonction de Hadwiger est une troisième piste, car elle semble mieux adaptée pour modéliser la distribution de la fécondité. L'inconvénient est qu'estimer ses paramètres peut prendre du temps et que leur interprétation n'est pas forcément évidente. Enfin, pourquoi ne pas proposer une fonction *ad hoc*, qui reflète fidèlement les données observées ? On peut alors être tenté d'estimer de façon non paramétrique la distribution des taux de fécondité, c'est-à-dire en

réalité en utilisant un très grand nombre de paramètres. La difficulté réside alors en la projection de ces très nombreux paramètres. On peut aussi penser, pour les trois composantes d'évolution de la population, à développer des modèles structurels, permettant d'expliquer l'évolution passée selon des mécanismes plus détaillés et reposant sur des variables externes, mais cela impose également d'avoir suffisamment d'éléments pour ensuite projeter l'évolution de ces variables. Par ailleurs, il serait très instructif de mener des études de sensibilité, qui permettraient de tester comment varie les résultats lorsque l'on modifie légèrement certaines hypothèses des modèles. Cela aiderait à mieux comprendre et quantifier le rôle précis de chaque composante dans l'évolution de la population.

On le voit, de nombreuses améliorations sont sans doute possibles. Elles demandent des investissements importants de recherche dans la compréhension et la modélisation des migrations, de la mortalité et de la fécondité, ce qui ne pourra qu'être bénéfique aux projections probabilistes de population, dont le degré d'incertitude dépend avant tout de nos connaissances (ou de nos ignorances) sur ces sujets. Enfin, il ne faudrait pas opposer les projections probabilistes de la population aux projections déterministes de la population. Ces dernières restent très utiles et permettent de tester ce qui se passerait à l'avenir, dans tel ou tel scénario. Les conclusions générales auxquelles on aboutit sont d'ailleurs très cohérentes avec celles des projections déterministes quant à l'évolution de la taille de la population et de sa structure par âge. Mais c'est avant tout aux utilisateurs des projections de population de choisir l'approche qui leur convient le mieux, selon leurs usages. Les projections probabilistes et déterministes sont deux manières différentes d'aborder l'incertain et d'essayer d'éclairer l'avenir. □

BIBLIOGRAPHIE

- Alkema, L., Raftery, A. E., Gerland, P., Clark, S. J., Pelletier, F. & Buettner, T. (2010).** Probabilistic projections of the total fertility rate for all countries. Center for statistics and the social sciences, University of Washington. *Working paper* N° 97. <https://doi.org/10.1007/s13524-011-0040-5>
- Bellamy, V. & Beaumel, C. (2016).** Bilan démographique 2015, le nombre de décès au plus haut depuis l'après-guerre. *Insee Première* N° 1581. <https://www.insee.fr/fr/statistiques/fichier/version-html/1908103/ip1581.pdf>
- Bijak, J., Alberts, I., Alho, J., Bryant, J., Buettner, T., Falkingham, J., Forster, J. J., Gerland, G., King, T., Onorante, L., Keilman, N., O'Hagan, A., Owens, D., Raftery, A. & Ševčíková, H. (2015).** Probabilistic population forecasts for informed decision-making. *Journal of Official Statistics*, 31(4), 537–544. <https://doi.org/10.1515/jos-2015-0033>

- Bijak, J. & Bryant, J. (2016).** Bayesian demography 250 years after Bayes. *Population Studies*, 70(1), 1–19. <https://doi.org/10.1080/00324728.2015.1122826>
- Blanpain, N. (2016).** Les hommes cadres vivent toujours 6 ans de plus que les hommes ouvriers. *Insee Première* N° 1584. <https://www.insee.fr/fr/statistiques/fichier/version-html/1908110/ip1584.pdf>
- Blanpain, N. (2020).** Is the Ageing of the French Population Unavoidable? *Economie et Statistique / Economics and Statistics*, ce numéro.
- Blanpain, N. & Buisson, G. (2016a).** Projections de population 2013-2070 pour la France : méthode et principaux résultats. Insee, *Document de travail* N° F1606. https://www.insee.fr/fr/statistiques/fichier/2400057/F1606_version_revue_septembre2018.pdf
- Blanpain, N. & Buisson, G. (2016b).** Projections de population à l'horizon 2070. *Insee Première* N° 1619. <https://www.insee.fr/fr/statistiques/fichier/version-html/2496228/ip1619.pdf>
- Booth, H. (2006).** Demographic forecasting: 1980 to 2005 in review. Australian National University, *Working Papers in Demography* N° 100. <https://demography.cass.anu.edu.au/sites/default/files/sod/publications/working-papers/100.pdf>
- Brutel, C. (2014).** Estimer les flux d'entrées sur le territoire à partir des enquêtes annuelles de recensement. Insee, *Document de travail* N° F1403. <https://www.insee.fr/fr/statistiques/fichier/1381143/F1403.pdf>
- Bryant, J. & Zhang, J. L. (2014).** Bayesian forecasting of demographics rates for small areas: emigration rates by age, sex and region in New Zealand. *Statistica Sinica Preprint* N° SS-14-200tR3. <https://www.jstor.org/stable/44114337>
- Conseil d'orientation des retraites (2017).** Évolutions et perspectives des retraites en France. Rapport annuel du COR. <https://www.cor-retraites.fr/sites/default/files/2019-06/doc-4055.pdf>
- Costemalle, V. (2015).** Projections de populations : l'ONU adopte une méthode bayésienne. *Statistique et Société*, 3(3), 9–14. <http://statistique-et-societe.fr/article/view/500/474>
- Dunstan, K. & Ball, C. (2016).** Demographic projections: user and producer experiences of adopting a stochastic approach. *Journal of Official Statistics*, 32(4), 947–962. <https://doi.org/10.1515/jos-2016-0050>
- Gerland, P., Raftery, A. E., Ševčíková, H., Li, N., Gu, D., Spoorenberg, T., Alkema, L., Fosdick, B. K., Chunn, J., Lalic, N., Bay, G., Buettner, T., Heilig, G. K. & Wilmoth, J. (2014).** World population stabilization unlikely this century. *Science*, 346(6206), 234–237. <https://doi.org/10.1126/science.1257469>
- Hyndman, R. J. & Booth, H. (2006).** Stochastic population forecasts using functional data models for mortality, fertility and migration. Australian National University, *Working Papers in Demography* N° 99. <https://demography.cass.anu.edu.au/sites/default/files/sod/publications/working-papers/99.pdf>
- Hyndman, R. J. & Ullah, S. (2007).** Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, 51(10), 4942–4956. <https://doi.org/10.1016/j.csda.2006.07.028>
- Kontis, V., Bennett, J. E., Mathers, C. D., Li, G., Foreman, K. & Ezzati, M. (2017).** Future life expectancy in 35 industrialised countries: projections with a Bayesian model ensemble. *The Lancet*, 389(10076), 1323–1335. [https://doi.org/10.1016/S0140-6736\(16\)32381-9](https://doi.org/10.1016/S0140-6736(16)32381-9)
- Lee, R. D. & Carter, L. R. (1992).** Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87(419), 659–671. <https://doi.org/10.2307/2290201>
- Lutz, W., Sanderson, W. & Scherbov, S. (2001).** The end of world population growth. *Nature*, 412, 543–545. <https://doi.org/10.1038/35087589>
- MacPherson, L. (2016).** National population projections : 2016(base)-2068. Statistics New Zealand. <https://www.stats.govt.nz/information-releases/national-population-projections-2016base2068>
- Organisation des Nations Unies, Département des affaires économiques et sociales, Division Population (2017).** World Population Prospects: The 2017 Revision, key findings and advance tables. *Working Paper* N° ESA/P/WP/248.
- Raftery, A. E., Chunn, J. L., Gerland, P. & Ševčíková, H. (2013).** Bayesian probabilistic projections of life expectancy for all countries. *Demography*, 50, 777–801. <https://doi.org/10.1007/s13524-012-0193-x>
- Raftery, A. E. (2014).** Use and communication of probabilistic forecasts. University of Washington. arXiv:1408.4812v1. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5403155/>
- Régnier-Loilier, A. & Vignoli, D. (2011).** Intentions de fécondité et obstacles à leur réalisation en France et en Italie. *Population-F*, 66(2), 401–432. <https://doi.org/10.3917/popu.1102.0401>
- Wiśniowski, A., Smith, P. W. F., Bijak, J., Raymer, J. & Forster, J. J. (2015).** Bayesian population forecasting: extending the Lee-Carter method. *Demography*, 52, 1035–1059. <https://doi.org/10.1007/s13524-015-0389-y>

