N° 520-521 - 2020

Economie et Statistique

Economics and Statistics

Varia

Dossier thématique

Projections de population

Articles

Accumulation de liquidités par les entreprises

Concentration et part du travail

Thematic Section

Population Projections

Articles

Accumulation of Cash by Firms

Concentration and Labour Share

Insee

Economie et Statistique

Economics and Statistics

Directeur de la publication / Director of Publication:
Jean-Luc TAVERNIER
Rédactrice en chef / Editor in Chief:
Sophie PONTHIEUX
Responsable éditorial / Editorial Manager: Pascal GODEFROY
Assistant éditorial / Editorial Assistant: …
Traductions / Translations: RWS Language Solutions
Chiltern Park, Chalfont St. Peter, Bucks, SL9 9FG Royaume-Uni
Maquette PAO et impression / CAP and printing: JOUVE
1, rue du Docteur-Sauvé, BP3, 53101 Mayenne

La revue est en accès libre sur le site **www.insee.fr**. Il est possible de s'abonner aux avis de parution sur le site. La revue peut être achetée sur le site **www.insee.fr**, rubrique « Services / Acheter nos publications ». La revue est également en vente dans 200 librairies à Paris et en province.

The journal is available in open access on the Insee website **www.insee.fr.** Publication alerts can be subscribed on-line. The printed version of the journal (in French) can be purchased on the Insee website **www.insee.fr**.

N° 520-521 - 2020

Economie et Statistique

Economics and Statistics

Economie et Statistique / Economics and Statistics

Issue 520-521 – 2020

VARIA

THEMATIC SECTION: POPULATION PROJECTIONS

ARTICLES

THEMATIC SECTION
Population Projections

Introduction

Laurent Toulemon*, Gilles Pison** and Isabelle Robert-Bobée***

While Insee is set to prepare new population projections for France, the thematic section featured in this issue is an opportunity to take stock of both the results of previous projections and the methods used in France and elsewhere.

Why do we make projections? Given that they present not only the current or past situation, but also the future, projections are a very specific category of scientific product. It is not until the future has become the present and then the past that we can assess their "accuracy" or their "errors", treating projections as forecasts. Indeed, they are rarely forecasts; projections are not all carried out with a view to coming true. For example, some projections present an undesirable future, in the aim of highlighting actions that could be taken to avoid it (anti-fulfilling prophecy) while others set an objective and explore ways to achieve it (pro-active projections). The projections presented in this section all start from a central scenario presented as the most likely future on the basis of current knowledge, and propose variants based on that scenario. When carrying out a projection, users take the central scenario as a forecast to shed light on present actions or suggest others.

How do we develop scenarios? The scenarios are most often developed by analysing past trends, over a longer or shorter period. Experts examine them and propose the best way to continue them. Sometimes, hypotheses can be excluded *a priori*, and in fact lead to limiting the range of possible evolutions. Given what has been observed in the past in France for example, it would be difficult to imagine considering scenarios of negative net migration (involving a greater number of people leaving the country than entering), or a fertility rate above the replacement level of 2.1 children per woman. Refining the hypotheses may be taken further, such as in the case of the mortality scenarios used in Insee's projections for France. As **Nathalie Blanpain** explains, these now take into account, for the first time, generational effects – the decrease in mortality rate stops for the generations born between 1940 and 1955, before resuming at a steady pace for those born afterwards, as observed over the last few decades.

How do we manage the uncertainty? Working from the central scenario, the projections offer alternative scenarios which allow us to take the uncertainty associated with the hypotheses into account. Since the end of the baby boom, all of Insee's projections have shown that population ageing, understood as being the increase in the proportion of elderly people, is unavoidable. This is a happy consequence of living longer, accelerated in France as the large generations of baby-boomers (born between 1946 and 1974) reach more advanced ages. The projections are therefore useful to forecast adaptations to the health system or retirement schemes. Comparing these with projections from neighbouring European countries also sheds light on France's future. By conducting a comparison with Eurostat projections for France and for the other EU Member States, we can learn a great deal. With the public dissemination of expert opinions, details on the methods and results,

** Ined ; ** Ined and Laboratoire d'Eco-anthropologie, UMR 7206 CNRS-MNHN-Université Paris Diderot ; *** Insee*

Translated from French

Citation: Toulemon, L., Pison, G. & Robert-Bobée, I. (2020). Introduction to the Thematic Section on Population Projections. *Economie et Statistique / Economics and Statistics*, 520-521, 5–7. https://doi.org/10.24187/ecostat.2020.520d.2028

Insee's projections for France form a tool that has taken on a central importance in the public debate in the country.

What is the benefit of probabilistic projections? The high and low scenarios proposed by Insee for France allow us to set the limits of the uncertainty, while the probabilistic projections incorporate the risk in different ways: there are no variants but rather a set of scenarios built on the basis of probability densities. The main advantage of these projections is being able to offer not only a central estimate but also a confidence interval for any derived indicator (for example, the proportion of women among the over 65s in 2070). **Vianney Costemalle** engages in this exercise for France. In addition to proving the feasibility of these projections by actually carrying them out, he shows some differences compared to the usual Insee projections. The central value of projected fertility for 2070 is the same (1.95 children per woman), but the uncertainty is higher: the 95% confidence interval, assimilated here to the gap between the high and low hypotheses, is [1.63; 2.26] compared with [1.8; 2.1] for the high and low scenarios. Conversely, the mortality scenarios are both more pessimistic and narrower: 88.4 years and 92.0 years for life expectancy at birth for men and women in 2070, plus or minus a year, compared with 90 and 93 years, plus or minus three years, in the high and low scenarios.

How do we evaluate the projections? One way of evaluating past projections consists in comparing them with actual developments. **Nico Keilman** has shown in previous research that, for 40 years, the projections have not come close to reality, concluding that we need to make probabilistic projections (Keilman, 2008). Here, he proposes a method for evaluating this type of projection, and applies it to those of three countries, France, Norway and the Netherlands. This allows him to revisit the projections he participated in 10 years ago and to show that they turned out to be more accurate than official projections, except in the case of France where the adjustments made in 1999 and 2006 were not correctly taken into account in the estimation of the parameters. He also shows that the errors are more marked for certain age groups, either because there is more uncertainty here or because the adjustments related specifically to those ages.

How do we build the projections? The components method used in the projections consists in estimating, for each year, net migration by sex and age, deaths by sex and age on the basis of the mortality rates, and the total number of births on the basis of the number of women of childbearing age and the fertility rates by age.

The method is very effective as the sex and age of the inhabitants are very easy to forecast: girls aged 10 in 2020 will become women aged 60 in 2070, if they are still alive. These very severe restrictions regarding sex and age enable us to develop population projections that are much more robust than other projections (for example, economic projections) and to propose long time horizons of at least 50 years. Yet other dimensions can also be taken into consideration: residential lifestyle for household projections (Jacquot, 2012), professional situation for labour force projections (Koubi & Marrakchi, 2017), health status for dependent population projections (Lecroart, 2013; Larbi & Roy, 2019); these are traditionally conducted by Insee or DREES by projecting the proportions and applying them to the results of the population projections. The projections can be more complex and dynamic, for example projections by area of residence for sub-national projections (Desrivierre, 2017), in which the rates of internal migration are used to determine the number of internal migrants, with overall consistency guaranteed as each exit from a region becomes an entry into another.

Calculating projections by taking into account dimensions other than sex and age? We could also include other dimensions in the projections. This is what **Anne Goujon** presents in her discussion of the difficulties involved in the exercise using projections based on level of education. She reviews the methods used for multi-state population projections and shows their potential added value (measure of human capital, feedback from education on fertility components, migrations, mortality). By way of example, she addresses other possible additional dimensions: diet, language spoken, political or religious opinions, and family network, and discusses the increased difficulty involved in the exercise when these different dimensions are included.

The UN World Population Prospects. The section begins with a presentation, by **Thomas Buettner**, of the most notable projection exercise to date: the United Nations World Population Prospects (WPP). First published just after the Second World War, in 1951, these projections are based on current population estimates and the desire to take a long-term view; the projections have been revised at regular intervals (currently every two years), with those published in 2019 comprising the 26th edition. The description of the components and their development at a continental level gives an idea of the work undertaken and progress made. The results and methods are now easily accessible and can be used as a reference by all other efforts in this field. The series of projections is very extensive, which allows us to compare the projections both with the actual developments in different countries or continents, or the entire world, and amongst themselves, with their developments resulting both from the revision, in each edition, of some of the past figures and modifications of future scenarios. Moving to probabilistic projections has, to some extent, allowed us to do away with high and low scenarios (±0.5 children per woman in all countries) and the confidence intervals used in probabilistic projections give rise to significant work in presenting the uncertainty and its limits when publishing the results.

References

Desrivierre, D. (2017). D'ici 2050, la population augmenterait dans toutes les régions de métropole. *Insee Première* N° 1652. https://www.insee.fr/fr/statistiques/2867738

Jacquot, A. (2012). La demande potentielle de logements à l'horizon 2030 : une estimation par la croissance attendue du nombre des ménages. Ministère de l'Environnement, de l'Énergie et de la Mer, *Le point sur* N° 135.
https://www.statistiques.developpement-durable.gouv.fr/sites/default/files/2018-10/LPS%20135.pdf

Keilman, N. (2008). European demographic forecasts have not become more accurate during the past 25 years. *Population and Development Review*, 34, 1, 137–153.
https://www.jstor.org/stable/25434662

Koubi, M. & Marrakchi, A. (2017). Projections à l'horizon 2070 - Une hausse moins soutenue du nombre d'actifs. *Insee Première* N° 1646. https://www.insee.fr/fr/statistiques/2844302

Larbi, K. & Roy, D. (2019). 4 millions de seniors seraient en perte d'autonomie en 2050. *Insee première* N° 1767. https://www.insee.fr/fr/statistiques/4196949

Lecroart, A. (2013). Personnes âgées dépendantes : projection à l'horizon 2060. DREES, *Dossier solidarités santé* N° 43.
https://drees.solidarites-sante.gouv.fr/etudes-et-statistiques/publications/les-dossiers-de-la-drees/dossiers-solidarite-et-sante/article/projection-des-populations-agees-dependantes-deux-methodes-d-estimation

World Population Prospects – A Long View

Thomas Buettner*

Abstract – There is no need to justify interest in population dynamics. But there is a pertinent need for sufficient, detailed and consistent evidence. Today, there is ample information about demographic trends for countries small, exceptionally large, and in-between. This was not always the case. Since the late 1940s, the United Nations Population Division endeavored to collect (often sparse) evidence for an increasingly complete picture known as World Population Prospects. Its evolution, through 26 revisions, is the topic of this article. It starts with the historical context, followed by brief discussions of the demographic components of change: fertility, mortality and (net) migration. Based on a reconstruction of past trends (or estimates), the Populations Division projects the population of today 235 countries or areas; the world's population could reach between 9.4 to 12.7 billion people, with a median of 10.9 billion. The article closes with suggestions about further improvements.

JEL Classification: J1, J13, F22, I1
Keywords: United Nations, population estimates, projections, fertility, mortality, migration, long-term trends

**Board member German Fund for World Population, former Assistant Director United Nations Population Divisions (planetbuettner@gmail.com)*

Acknowledgments – The author thanks two anonymous reviewers.

Received September 2019, accepted July 2020.

Citation: Buettner, T. (2020). World Population Prospects – A Long View. *Economie et Statistique / Economics and Statistics*, 520-521, 9–27. https://doi.org/10.24187/ecostat.2020.520d.2030

Prologue

It was the worst of time. It was 1944, WWII still raged on and would continue for about another year. It was a time of measured hope, too. In August 1944 (19.8.1944), the French Resistance begins an uprising in Paris and by August 25, Paris is liberated. On September 3, French and American troops liberated Lyon,[1] *and the British liberated Brussels. The Red Army had reached Warsaw, the Pacific Theater saw heavy fighting in the battle of the Philippines.*

At this ominous time, the University of Chicago held the Twentieth Institute of the Norman Wait Harris Foundation (September 4-8, 1944), with the general theme "Food in International Relations." Frank Notestein, the director of the Office of Population Research at Princeton University, presented a paper entitled "Population – The Long View" that reviewed the global population trends of the past three centuries and laid out the contours of a conceptual framework for global population projections (Notestein, 1945). He identified the population growth of his time as transitional and distinguished three demographic types, or regimes, that represent different stages of that growth (incipient decline, transitional growth, and high growth potential). He also stated the demographic cause of the transitional growth – mortality decline followed by fertility decline later – and posited that (rapid) aging of populations is unavoidable. He concludes: "[...] it appears [...] that sensible planning for the future should be based on the assumption that the world will have at least 3 billion people by the year 2000. [...] Food production will have to increase much more rapidly than population, and equally swift developments must occur in the fields of industrial production, education, public health, and government. For it is only when rising levels of living, improved health, increasing education, and rising hope for the future give new value and dignity to the individual life that old customs break, and fertility comes under control. [...] In the long run it remains true that the control of mortality without the control of fertility is impossible." (Notestein, 1945, p. 57). His "short" summary (just 21 pages!) provided a lasting foundation for social analysis beyond demography. Notestein's attempt in that article to estimate the world population in 2000, however, failed spectacularly. The world population in 2000 was not 3, but 6 billion.

On 24 October 1945, the United Nations was established. Less than a year later, its Economic and Social Council, on 3 October 1946, created the Population Commission[2] *to afford "advice and assistance on matters affecting or affected by population change". At the same time, the Population Division as the Commissions Secretariat was created, with Frank Notestein its first Director.*

The need for population projections was realized early. The second session of the Population Commission considered the need for population estimates and forecasts and decided to set priorities: "Noting that the requirements for such estimates and forecasts were extremely large, the Commission (E/571) adopted a scheme of priorities designed to make available as soon as possible the estimates and forecasts which were most essential for the work of the various organs of the United Nations. The Commission considered that the first object should be to compile current estimates of the total population, as of a uniform, recent date, for all countries of the world. Other data which it recommended should have a high priority were estimates of population by sex and age groups for recent dates, forecasts of total population and sex and age groups for dates in the near future (1948, 1949 and 1950), and longer-range forecasts." (Population Commission, 1947, p. 20; United Nations, 1948, p. 640).

The notion of world population is not an invention of the 20th century. But it was the 20th century that began to measure it in earnest and in detail, its historic evolution first (Biraben, 1979, 2006; Durand, 1974). From the scattered empirical records of historical population growth emerged evidence that human populations by no means were destined to grow in an exponential fashion (the geometric growth envisioned by Malthus). But then, what would the future hold? Some theoretical propositions came from a synthesis of the empirical past in some countries, called the demographic revolution (Landry, 1934) or, later, the demographic transition (Davis, 1945; Notestein, 1945). It formulated a concept of a universal process from high to low levels of fertility and mortality. It has guided the demographers at the Population Division well, most of the time.

1. *Leon Tabah, the fifth director of the United Nations Population Division (from 1972 to 1984), took part in the liberation of Lyon and was awarded the* Médaille de la Résistance.
2. *Renamed, in 1994, the Commission on Population and Development pursuant to A/RES/49/128, para. 24, of 19 December 1994.*

The World Population Prospects (WPP) have been work in progress from its beginning. As new demographic and other related data became available, as methodological improvements were developed, and as computational tools became more powerful, existing estimates and projections were revised, updated, and expanded, and revised again. Formats and titles of the outputs also changed, as did geographic coverage, demographic detail, and projection horizons. In many respects, comparing current revision of WPP with earlier ones is meeting considerable challenges. One of the most significant changes are geographic and political settings. Countries were gaining independences, changed their names, some countries split, notably the Soviet Union into 15 successor states, others united or re-united.

The evolution of the estimates and projections over time – now at its 26 iteration or revision – is a testament to the commitment of the international community and the dedication of the staff of the Population Division to its original mandate dating back to the 1940s.

In its current form, the WPP is an impressive account of demographic change for all 235 countries[3] of the world for the past 70 years, from 1950 through 2020.[4] This account is not just a collection of relevant demographic indicators: it has evolved into a complete and internally consistent reconstruction of the world's demographic history. It contains demographic detail that must have seem impossible for the demographers that started the project. The more visible part of the project – the population projections – are now available up to the end of the century with equal detail. Recently, projections results are produced with prediction intervals, plus certain illustrative scenarios.

This paper is a brief review of the history of efforts, approaches, failures, and successes of the United Nations WPP. The evolution of projection methodology will here not receive the deserved attention due to space limitations.[5] Population projections are here primarily understood as a powerful instrument of analyzing and understanding current conditions (Keyfitz, 1972), including our current understanding of future trends in fertility, mortality and migration.[6] The reference to the current conditions and understanding necessarily imply that these projections are an ongoing process. Here, we take the 2019 Revision as the reference, assuming it provides the best summary of past demographic trends (1950-2020), and projections (2020-2100). Future revisions of WPP will certainly introduce more changes, both for past estimates and for the projections: WPP will continue be work in progress.

The story of the United Nations population projections[7] did not begin with projecting the future, but with the past. In 1949, the study *World Population Trends 1920-1947* was published, presenting "[…] estimates of population, birth and death rates, life expectancy and age structure of the population, for the world and its principal regions" (United Nations, 1949, p. iii).

When the demographers at the United Nations Population Division published their first projections for the world in 1951 (United Nations, 1951), they based it on partial times series up to 1950, defining that year as the base year. [8] From that time on, the year 1950 marked the start of WPP. Successive revisions kept the year 1950 as the beginning of the exercise but moved the base year forward to the calendar year (that was divisible by five) nearest to the year the revision was completed (see Online Appendix C2). When new data from censuses, vital statistics, surveys, and other sources became available, the projected population for the new base year from the previous revision had to be updated. Obviously, new and updated base year population estimates would impact the outcomes of the projection exercises. But updating the base population estimates did also have an impact on the past: if the demographic accounting identity was to be maintained for the whole projection exercise, past population estimates, and the associated demographic variables, had to be revised, too. In other words: not only future populations were a moving target, but past estimates, too.

The rich history of the WPP may be presented in many ways. The usual presentation is often focused on the population – its size, composition, and geographical distribution. The driving forces of demographic change – fertility, mortality, and migration – are often less prominently addressed. Here, they are presented first. Then we look at population estimates in Section 2.

3. *As is the usual practice of WPP, the term "country" as used in this text also refers, as appropriate, to territories or areas. A more detailed classification is listed in Online Appendix C1 to this article. Link to the Online Appendices at the end of the article.*
4. *The latest revision of the United Nations' World Population Prospects, released in 2019, chose the year 2020 as the base year of its projections. The data for the year 2020 is, of course, projections, based on data available through 2019.*
5. *Online Appendix C2 lists some of the changes to the projection methodology. For the latest version, see United Nations (2019a, 2019b).*
6. *Caswell stated: "Population projections reveal something about present conditions […], not about the future behavior of the population" (Caswell, 2001, p. 30).*
7. *For more detailed account of the history of the past 26 revisions of the WPP see Online Appendix C2.*
8. *Sometimes called jump-off year or launch year.*

1. The Evolution of Estimates and Projections: Components

1.1. Fertility

The past 70 year have seen sustained, sometimes dramatic, reductions in fertility. On average, the number of children per woman fell from 5.0 to 2.5 children for the world between 1950 and 2020, or about 0.2 children per woman per quinquennium. Such an average does not show the vast differences between countries, regions, and subregions during that period. A comparison of regions (Table 1, Figure I) shows an onset of fertility decline later than 1950, or even a temporary increase (Northern America, Oceania). It was not before the decade of the 1960s that sustained decline of fertility in most of these large aggregates was manifesting itself. Africa, with the highest average fertility in 1950-1955 of 6.6 children per woman, entered the fertility transition, on average, not before the decade of the 1970s. In 2015-2020 (base period), Africa has still the highest total fertility of 4.4 children. All other regions have transitioned to low fertility around or even well below the replacement level (Asia, Europe, Latin America, Northern America).

Average fertility levels for large aggregates or the world mask the existing great variations for the 235 countries. Currently, that is 2015-2020, fertility ranges from 7.0 children per woman (Niger) to 1.1 children per woman (Republic of Korea).

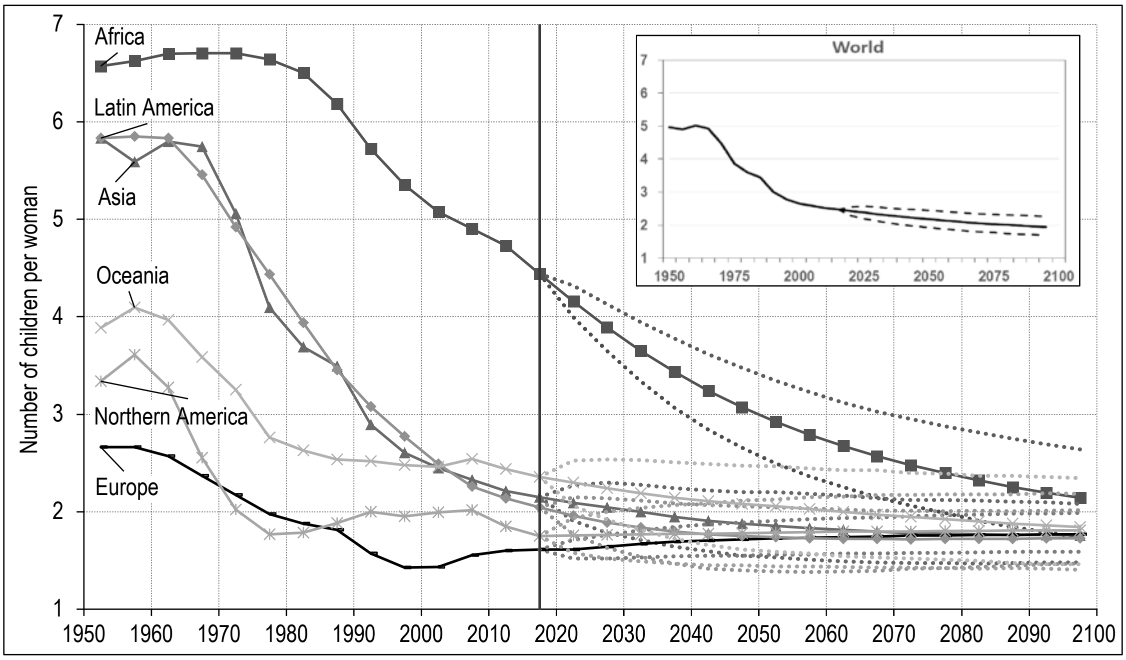
Remarkably, close to half (49%) of humankind lives already in countries with fertility at or below the replacement level of 2.1 children per woman (Table 2). Intermediate fertility, that is fertility between 2.1 and 5 children per woman, is estimated for another 46% of the world's

Table 1 – **Total fertility estimates and projections by regions, 1950-2100**

	Number of children per woman						Change (%)	
	1950-1955	1975-1980	1995-2000	2015-2020	2045-2050	2095-2100	1950-2020	2015-2100
World	5.0	3.9	2.8	2.5	2.2	1.9	-50	-22
Africa	6.6	6.6	5.4	4.4	3.1	2.1	-32	-52
Asia	5.8	4.1	2.6	2.2	1.9	1.8	-63	-18
Europe	2.7	2.0	1.4	1.6	1.7	1.8	-40	+10
Latin America	5.8	4.4	2.8	2.0	1.8	1.7	-65	-15
Northern America	3.3	1.8	2.0	1.8	1.8	1.8	-47	+3
Oceania	3.9	2.8	2.5	2.4	2.1	1.8	-39	-22

Sources: WPP 2019.

Figure I – **Total fertility estimates and projections by regions, 1950-2100**



Notes: Solid lines for the median of the prediction interval, dotted lines for the 95% prediction interval (upper, lower).
Sources: WPP 2019.

Table 2 – **Number of countries by fertility level and their share of world population, 1950-2020**

Births per woman	Number of countries			% of world population		
	1950-1955	1980-1985	2015-2020	1950	1980	2015
Below replacement (less than 2.1)	6	50	115	0.4	22.8	52.1
Intermediate (2.1 to 5)	68	94	109	34.1	60.1	43.0
High (more than 5)	161	91	11	65.5	17.1	4.9
Total	235	235	235	100	100	100

Sources: WPP 2019, author's calculations.

population. High fertility of 5 and more children is estimated to occur in about 5% of the world's population, that is in 11 countries. All these countries with high fertility level are found in Africa. The largest such countries are Nigeria, the Democratic Republic of Congo, the United Republic of Tanzania, and Uganda.

The assumptions regarding future fertility trends have the largest impact of population trends. Compared with mortality and migration, fertility is the main driver of population change and has been a major focus of policy interventions. It is also one of the items for which data from "statistically underdeveloped areas" have become available relatively soon and regularly.[9]

A detailed analysis of past fertility assumptions and their adjustments in subsequent revisions is outside the scope of this article (but would be interesting). Some demographers have criticized the United Nations for assuming for a long time an ultimate convergence of fertility to replacement fertility (2.1 children model) for low fertility countries. While the initial transition theory at least provided guidance about the direction of fertility to lower levels, the post transition situation does not profit from such guiding idea. The situation regarding the post-transition fertility is similar (but not equal) to the first demographic transition: the onset of fertility decline is the most uncertain factor in the first, and the level of completed fertility is the most uncertain in the second, if such level then exists. Once fertility decline started, the first demographic transition pointed to the direction of declining fertility and thus informed population projections relatively reliably. The second demographic transition, once (very) low fertility is reached, does not make strong arguments for an ultimate level of fertility, if any.

Vallin & Caselli noted that global population projections would become less reliable as most countries passed through the demographic transition: "In an era in which the great historical change, called the demographic transition, is coming to an end, the paradigm of the same name is of no assistance in predicting what will follow." (Vallin & Caselli, 2006, p. 231).

The introduction of a Bayesian hierarchical model to predict fertility levels based on past trends is an attempt to handle, inter alia, this conceptual uncertainty. It introduces an ultimate (low) fertility level that is no longer uniform at replacement level, but may be much lower, at different levels and arrived at in different times in the future. Once that stage is reached, fertility stays constant. But even this assumption of decrements that are approaching zero is accompanied by model-generated uncertainty bounds, or prediction intervals.

The 2019 Revision assumed a continued but varied fertility transition for countries with above replacement fertility (see Table 1). In terms of the aggregate level of regions, Africa, Asia, Latin America, and Oceania are expected to experience a long-term fertility reduction until 2100 (medium variant). Europe and Northern America could see a slight recovery of their low fertility level in the long run. By the end of the projection horizon all regions could be at or even below replacement level fertility.

At the end of the projection horizon in 2100 the number of countries with fertility levels above replacement levels is projected to shrink from 124 in 2015-2020 to only 21 at the end of the projection horizon in 2095-2100 (Table 3). All the remaining 21 countries would have rather moderate fertility levels, none higher than 2.5 children per woman.

1.2. Mortality

During the past 70 years the countries of the world and its regions have experienced a remarkable success in reducing mortality, by eliminating or controlling certain infectious diseases, stabilizing, and improving health care and improving overall

9. Starting in 1984, the Demographic and Health Surveys (DHS) have become an indispensable source of demographic information with more than 300 surveys in more than 90 developing countries. Methodologically similar and equally important, the Multiple Indicator Cluster Surveys (MICS), implemented by UNICEF, has collected a multitude of demographic and other data in 358 surveys in 118 countries.

Table 3 – **Number of countries by fertility level and their share of world population, 2015-2100**

Births per woman	Number of countries			% of world population		
	2015-2020	2045-2050	2095-2100	2015	2045	2095
Below replacement (less than 2.1)	121	164	214	52.1	71.1	79.5
Intermediate (2.1 to 5)	106	71	21	43.0	28.9	20.5
High (more than 5)	8	0	0	4.9	0.0	0.0
Total	235	235	235	100	100	100

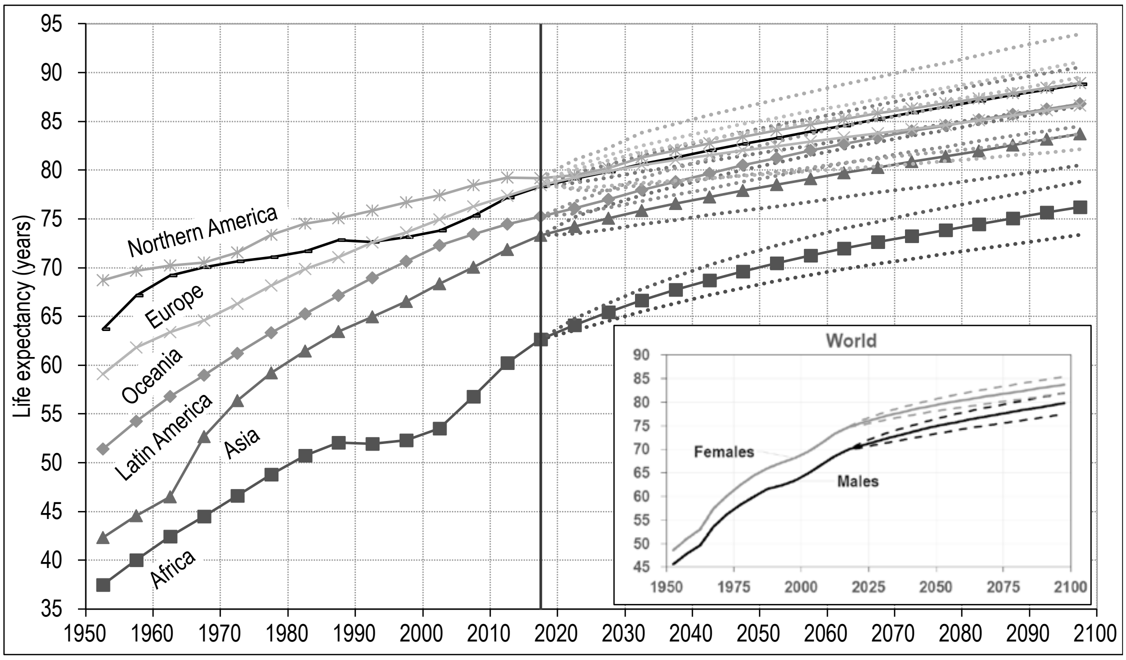
Sources: WPP 2019, author's calculations.

living conditions. This progress was not steady or without backslashes, but it happened. The 2019 Revision documents this transition to lower mortality (Figure II). Mortality declined for all countries and for almost all quinquennial periods for both males and females. There were some temporal exceptions to this global trend, caused, for some countries, by natural disasters, famine, civil strife, regional military conflicts and, notably, the HIV/AIDS pandemic. For the large aggregates of regions such temporary trend reversals are barely visible, except for Africa and Europe.[10]

Between 1950 and 2020, life expectancy for the world increased, on average and for both sexes combined, by about 25 years (Table 4).

10. The stall in life expectancy in Europe between 1985 and 2000 is mainly driven by increasing mortality in many successor states of the former Soviet Union as well as former Yugoslavia.

Figure II – **Life expectancy estimates for both sexes combined by regions, 1950-2100**



Notes: Solid lines for the median of the prediction interval, dotted lines for the 95% prediction interval (upper, lower).
Sources: WPP 2019.

Table 4 – **Life expectancy for both sexes combined by regions, 1950-2100**

	Life expectancy at birth (years)						Change (years)	
	1950-1955	1975-1980	1995-2000	2015-2020	2045-2050	2095-2100	1950-2020	2015-2100
World	47.0	60.3	65.6	72.3	76.8	81.7	25.3	9.4
Africa	37.5	48.8	52.3	62.7	69.6	76.2	25.2	13.6
Asia	42.3	59.2	66.6	73.3	77.9	83.7	31.0	10.5
Europe	63.7	71.1	73.1	78.3	82.7	88.8	14.6	10.5
Latin America	51.4	63.3	70.7	75.2	80.5	86.8	23.8	11.6
Northern America	68.7	73.3	76.7	79.2	83.4	88.9	10.4	9.8
Oceania	59.1	68.2	73.6	78.4	82.0	86.6	19.3	8.2

Sources: WPP 2019.

The largest absolute increase was observed for Asia, with 31 years over 70 years, followed by Africa gaining 25.2 years of lifetime per person on average. The smallest increase between 1950 and 2020 was estimated for Northern America (10.4 years), which had the highest level of life expectancy in 1950-1955. Most regions retained their relative position (except for a relatively minor crossover between Europe and Oceania). Clearly visible is also the levelling-off of life expectancy in Africa between 1985 and 2000, caused mostly by the HIV/AIDS epidemic (see below).

All countries participated in that impressive reduction in mortality, but at quite different times and with different paces. The number of countries with exceedingly high mortality (and corresponding low life expectancy) of less than 45 years dropped from 80 in 1950-1955 to just 10 thirty years later in 1980-1985, and by 2015-2020, no country was found at this level. At the same time, the number of countries with life expectancy above 75 years, increased to 133 in 2015-2020, while in 1950-1955, not one country was in that category (Table 5).

The remarkable reduction in overall mortality was caused to a large part by a dramatic decline in infant and child mortality. In 1950-1955, about one out of five newborn children did not reach its 5th birthday. Even in Europe, the under-five mortality rate was about one out of 10. In Africa and Asia, child mortality was exceedingly high: More than one in three children in Africa did not experience their 5th birthday; in Asia it was about one in four. That changed dramatically over the course of the 70 years that followed. Africa's under-five mortality rate was, in 2015-2020, where Europe stood during 1950-1955, Asia's under-five mortality rate resembles today that of Northern America in 1950-1955 (Table 6). The trend of significant reductions in child mortality is expected to continue over the projection horizon to very low levels. All estimated mortality data document a transition, still underway in many parts of the world, from early to late deaths. If low levels of child mortality subsist, early mortality has no major impact on the projections, except for Africa.

Progress is not destiny. One example of an unexpected severe reversal of mortality trends was the HIV/AIDS pandemic. Modeling the mortality impact of the HIV/AIDS epidemic started with the 1992 Revision and turned out to be a tremendous challenge. Limited empirical evidence had to be transformed into indicators of the epidemic (prevalence, incidence estimates) and further into age-specific mortality schedules for the affected populations. A competing risk model combined the mortality of the infected and the not infected population into a general dynamic mortality pattern. An example of the substantial uncertainty of the measurements

Table 5 – **Number of countries by mortality level and their share of world population, 1950-2020**

Life expectancy at birth, both sexes combined	Number of countries			% of world population		
	1950-1955	1980-1985	2015-2020	1950	1980	2015
<45	82	11	0	57	2	0
45-55	48	40	5	8	26	3
55-65	72	50	35	17	15	9
65-75	33	119	81	18	52	42
75+	0	15	114	0	5	46
Total	235	235	235	100	100	100

Sources: WPP 2019.

Table 6 – **Under-five mortality by regions, 1950-2100**

	Under-five mortality (deaths under age five per 1,000 live births)					
	1950-1955	1975-1980	1995-2000	2015-2020	2045-2050	2095-2100
World	213	124	82	40	22	12
Africa	311	200	151	71	36	18
Asia	234	127	73	31	15	6
Europe	93	26	12	5	2	1
Latin America	187	92	38	19	9	4
Northern America	36	17	9	7	4	2
Oceania	94	49	34	23	12	5

Sources: WPP 2019.

and models is seen in the comparison of life expectancy trends for Zimbabwe from the 1992[11] to the 2019 Revision of WPP (Figure III).

There was initially a clear underestimation, followed by an overestimation of the epidemic's mortality impact. Comparing the largest impact on life expectancy for Zimbabwe across several revisions, the 1992 Revision projected a life expectancy of 57.3 years, while the 2002 Revision came out with extreme low 33.1 years. The 2019 Revision, which relied on a much better empirical basis and revised epidemiological models, estimated life expectancy for 2000-2005 at 43.7 years for both sexes combined. The large variability of life expectancy estimates and projections, which has affected other countries with a significant burden of the epidemic as well, is clearly due to extremely limited empirical evidence initially about the dynamics of HIV/AIDS incidence and prevalence. All revisions, though, assumed and expected the Epidemic to be a temporary phenomenon, which is manifested by the ultimate rise of life expectancy for all revisions since 1992. This was in the beginning a matter of (institutional) optimism but was later confirmed after increasingly effective drugs became available, plus better testing und information.

Note that the experience of WPP modeling HIV/AIDS shows that emerging issues require a patient and repeated revisiting the issue. One-time estimates are often of limited validity. In this respect, the Population Division is well prepared to the continuous observation, estimation, and evaluation of such phenomena[12] thanks to its institutional stability.

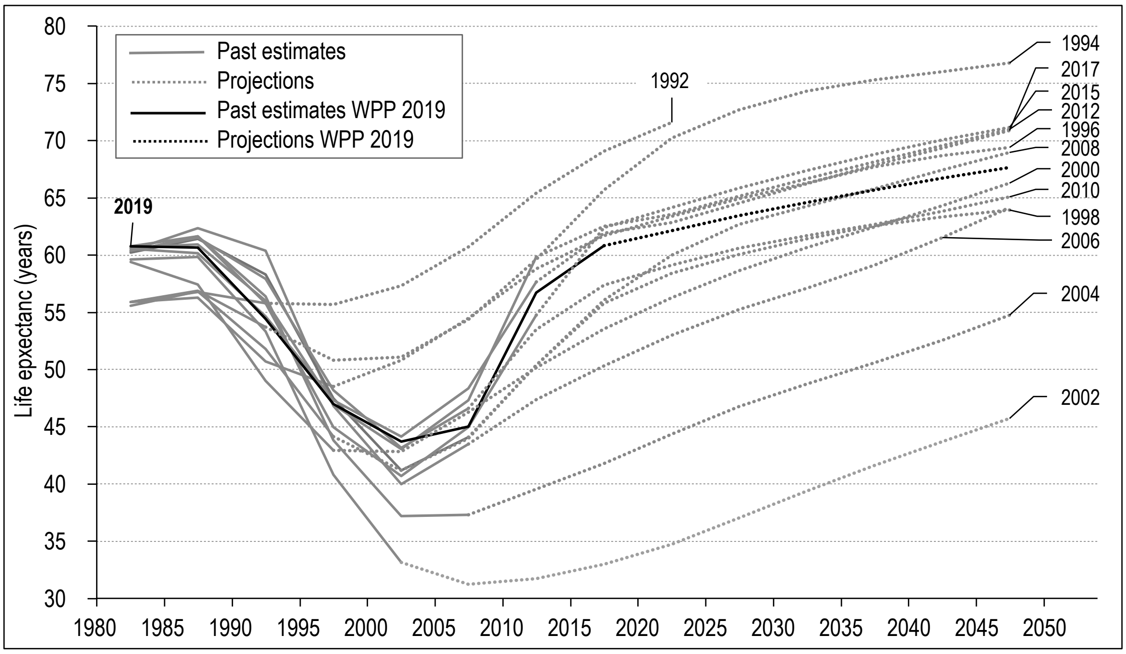
Is the transition to lower mortality expected to continue, and to which levels? The demographers at the UN provide tentative answers in their projections. Assuming, as usual, due progress in the future (here: progress against mortality), life expectancy for both sexes is projected to rise for the world as a whole (Figure II, Table 7): by 2095-2100, no country would have less than 65 years of life expectancy, and a majority of 211 countries even more than 75 years.

Global mortality projections for all countries of the world would not have been possible without models of mortality change and age patterns of mortality, predominantly based on historical data from developed countries. This reliance on models for many countries was and still is necessitated by a dramatic gap in registering mortality events, especially adult mortality, in developing countries. In 2007, a series of WHO analyses found almost no progress between 1970 and 2004 in covering adult mortality, especially in developing countries (AbouZahr *et al.*, 2007, 2015; Mikkelsen *et al.*, 2015; Setel *et al.*, 2007).

11. The United Nations Population Division incorporated the impact of HIV/AIDS since the 1992 Revision, using information from WHO's Global Programme on AIDS, and subsequently from the Joint United Nations Programme on HIV and AIDS (UNAIDS), which was formed in July 1994.

12. This is, of course, also true for UNAIDS providing continuous awareness of the HIV/AIDS pandemic.

Figure III – **Life expectancy estimates and projections for Zimbabwe since 1980 by revision**



Sources: WPP 1992 through 2019.

Table 7 – **Number of countries by mortality level and their share of world population, 2015-2100**

Life expectancy at birth, both sexes combined	Number of countries			% of world population		
	2015-2020	2045-2050	2095-2100	2015	2045	2095
<45	0	0	0	0	0	0
45-55	5	0	0	3	0	0
55-65	35	9	0	9	5	0
65-75	81	59	24	42	41	17
75+	114	167	211	46	53	83
Total	235	235	235	100	100	100

Sources: WPP 2019.

The empirical evidence of mortality trends has somewhat improved in most developing countries, mostly by sample surveys, not civil registration. Much efforts have been spent to distill the best estimates from the various sources available, but the evidence base remains shaky. It is therefore a great improvement that life expectancy projections are now showing ranges of uncertainty. This is even more remarkable as all revisions before the 2012 employed only one central variant.

For the projection period, the 2019 Revision assumes a steady increase in life expectancy, but at a decelerating pace (see Table 4). In 2095-2100, life expectancy for the world would reach about 82 years, an increase by more than 9 years. The largest increase is projected for the region with the highest mortality levels at the base period of 2015-2020: Africa's average life expectancy is expected to increase by almost 14 years, from 62.7 to 76.2 years. Regions with lower mortality in the base period are projected to gain smaller amount but retain their leading positions.

1.3. International Migration

International migration is the most challenging element in demographic accounting. Even countries with a well-developed statistical system often do not register international migration sufficiently, consistently, and reliably. Reasons for this are many. A prominent one is that countries rely on different definitions and procedures for what constitutes a migration event and who is to be registered as migrant. Therefore, statistics on international migration are often internationally not compatible. International migration is, therefore, often reduced to a residual measure. Yet, international migration as a flow of people involves at least two countries. International trade, e.g. flows of goods, is better documented than the movement of people.

Due to the lack of sufficiently complete and reliable migration flow data, WPP used net migration estimates and projections. Net migration is a complicated thing, as there is no real living "net migrant". It is better understood as a residual measure necessary to "close" the balance identity of demography, which is always also in danger of attracting the measurement errors of censuses, or births and deaths registration. In contrast, international migration flows, unlike net migration, are affecting both origin and destination countries. Thus, net migration is void of a critical aspect of international migration – the origin-destination link. Hence, it is spatially ignorant and relevant only for the country concerned.[13] It is also prone to exhibit unusual or inconceivable age patterns.

For the past, WPP show consistent and sustained geographical divisions regarding net migration of regions gaining and those region loosing people through migration. Since 1950, Europe has gained about 43 million people by 2020, Northern America 64 million people, and Oceania about almost 8 million people. At the same time, Africa lost 28 million people, Asia 44 million people, and Latin America about 43 million people (Table 8). These overall figures are significant, but not dramatic. After all, migration (net migration) is but a small component of population changes at the aggregate level. For individual countries and certain time periods, migration may play a considerable and critical role, however.

13. Because net migration is spatially ignorant, it does not balance automatically at the world level. Even if major migration flows were taken into consideration when estimating net migration estimates, a separate step of balancing the migration component is necessary to sum the migration component to zero for the world.

Table 8 – **Net migration estimates by regions, 1950-2020**

	Net migration (million)			
	1950-1980	1980-2000	2000-2020	1950-2020
World	0.0	0.0	0.0	0.0
Africa	-7.2	-7.7	-12.8	-27.7
Asia	+1.2	-12.4	-32.9	-44.1
Europe	-0.9	+12.1	+31.5	+42.7
Latin America	-11.6	-16.4	-14.6	-42.6
Northern America	+16.1	+22.8	+25.3	+64.2
Oceania	+2.4	+1.6	+3.4	+7.5

Sources: WPP 2019.

Expressing net migration intensity as Crude Net Migration Rate, that is the amount of net migration per 1,000 population,[14] reveals its relative small impact on demographic dynamics (Table 9).

For geographic regions, the net migration rate has been highest for those gaining population through net migration (net immigration) – Northern America and Oceania, followed by Europe. In comparison, the negative net migration rates – indicating population loss – are much smaller – below 1 per one thousand population.

Nevertheless, international migration is becoming increasingly important for population dynamics, especially in settings of low or very low fertility and the ensuing population ageing and eventually even population decline. In addition, international migration is also a factor of eminent political importance.

What are the prospects of future migration trends according to the WPP? Not surprisingly, assumptions for future migration are still starkly reflecting the lack of data, theories, even clear trends. Consequently, migration assumptions have been therefore quite simple.

The 2019 Revision has changed the assumption from a diminishing long-term trend to assuming a constant amount of net migration throughout most of the projection period. Figure IV shows the aggregate levels and Figure V the rates of net migration for the six geographic regions of the world. The picture is one of stasis, without temporal variations. The gain through net migration (Table 10) is largest for North America (105 million) and for Europe (64 million), while

14. The net migration rate is the average per quinquennium.

Table 9 – **Net migration rate estimates by regions, 1950-2020**

	Net migration per 1,000 population			
	1950-1955	1975-1980	1995-2000	2015-2020
World	0.00	0.00	0.00	0.00
Africa	-0.51	-0.72	-0.65	-0.37
Asia	0.14	-0.10	-0.33	-0.38
Europe	-0.62	0.59	0.87	1.83
Latin America	-0.52	-2.06	-1.86	-0.82
Northern America	1.58	3.38	6.35	3.30
Oceania	6.13	0.82	2.03	3.79

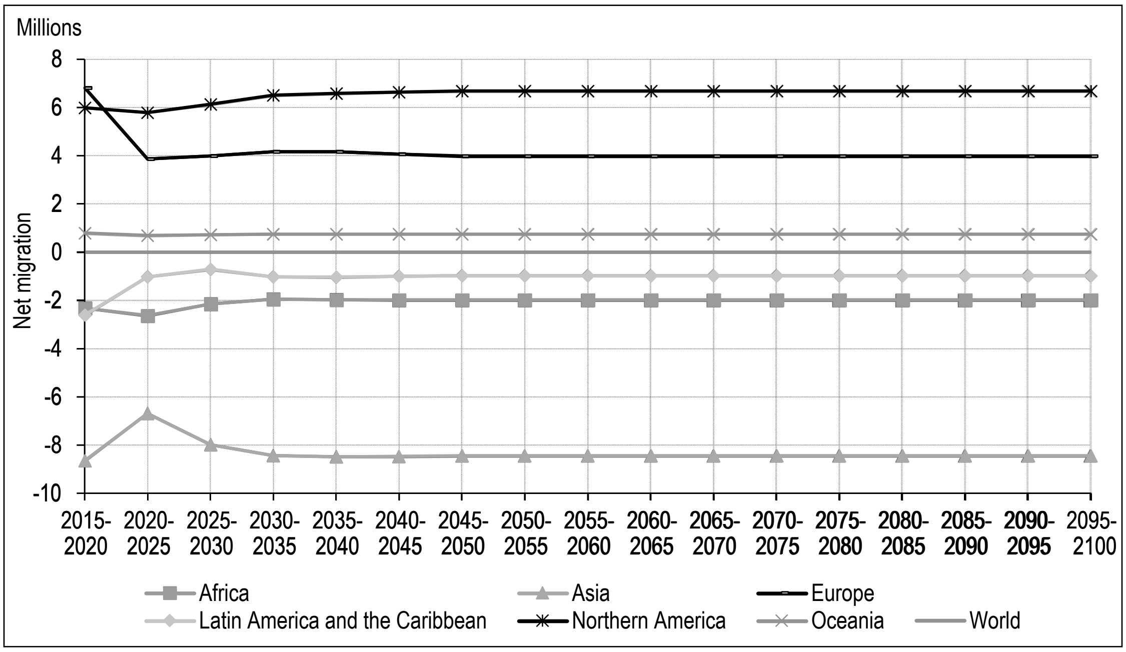
Sources: WPP 2019.

Table 10 – **Net migration projections by regions, 2020-2100**

	Net migration (million)		
	2020-2050	2050-2100	2020-2100
World	0.0	0.0	0.0
Africa	-12.7	-19.8	-32.5
Asia	-48.5	-84.4	-132.9
Europe	24.2	39.8	64.0
Latin America	-5.8	-9.8	-15.6
Northern America	38.3	66.7	105.0
Oceania	4.4	7.5	11.9

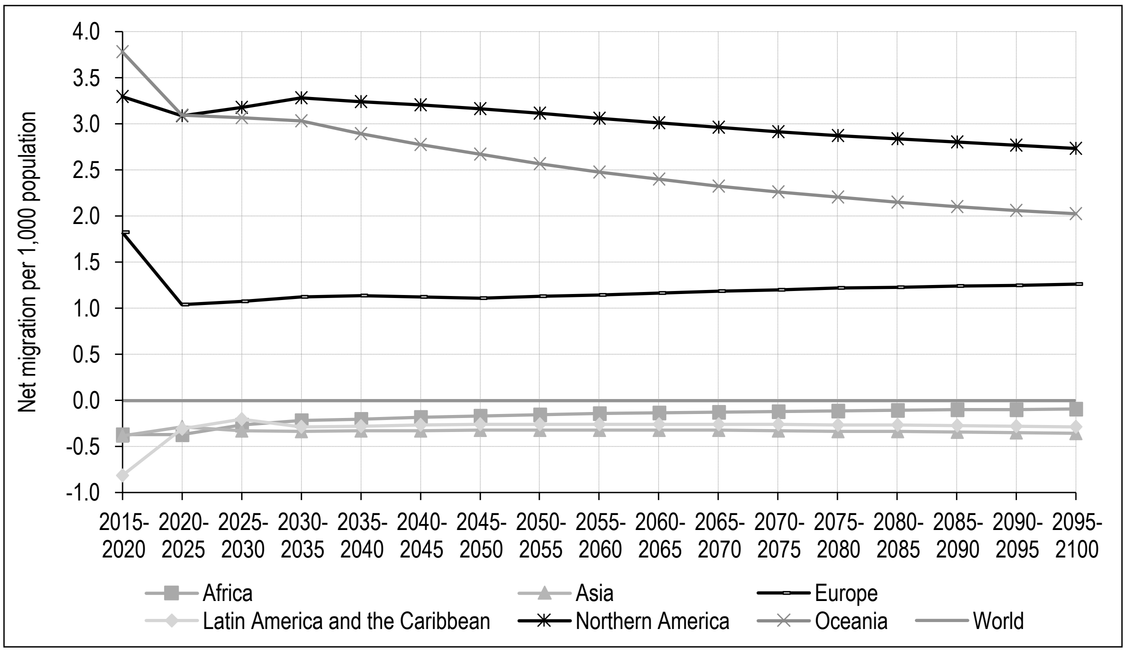
Sources: WPP 2019.

Figure IV – **Net migration projections by regions, 2015-2100**



Sources: WPP 2019.

Figure V – **Net migration rates by regions, 2015-2100**



Sources: WPP 2019.

Asia is the region with the highest loss through net migration (133 million).

In relative terms, e.g. as net migration per 1,000 population, the figures show changes caused by population dynamics: increasing intensities for declining populations and decreasing intensities for growing populations.

2. Population

Here it all comes together. Combining the assumptions about future fertility, mortality, migration, and the base population by using the cohort-component projection method[15] produces a consistent and detailed picture of the demographic future for each country.[16] Because the migration component operating as net migration is spatially ignorant, consistency at the aggregate level is not automatic. Therefore, after all countries of the world are projected, a second step of consolidation is often required to ensure a migration balance of zero for the world.[17]

Over the course of the 26 revisions of WPP, past population estimates were always an integral part of it, but in different levels of completeness, detail, and consistency. For many past revisions, past estimates were somewhat restricted to a reduced set of indicators: populations by age and sex at quinquennial dates and select indicators for the demographic components of change for quinquennial periods. Internal consistency was not ensured, only a full treatment with a cohort-component approach would do that. Step-by-step, the demographers moved the base year of the estimates and projections back to 1950 for ensuring consistency between the components of change and the population figures even for each age group and by sex. By the 2012 Revision, that process was finished, producing a full account of past demographic trends. The difference between past estimates and projections is now only that the former has only one variant, while the latter has several. It has been noted that the process of establishing the past and producing the best estimates of the base populations may well be the most laborious and time-consuming part of the whole exercise.[18]

The rich history of past population estimates between 1950 and 2020 is beyond the scope of the paper. Instead, we focus on the slow iteration of past world population figures to the most recent ones by past revisions. We calculated the relative difference between estimates and projections for certain calendar years – 1950, 1980,

15. The WPP have used the cohort-component method for most of its revisions but used simpler methods before the 1963 Revision. For a more detailed account of methods and assumptions see Timeline in Online Appendix C2.

16. Recall that technically the projections start in 1950, not 2020. In other words, projections are used for the demographic reconstruction of the past 1950-2020 and called past estimates, and the term projection is retained for the «true» projection period, here from 2020-2100 (characterized by different projection variants and, in some cased with prediction intervals).

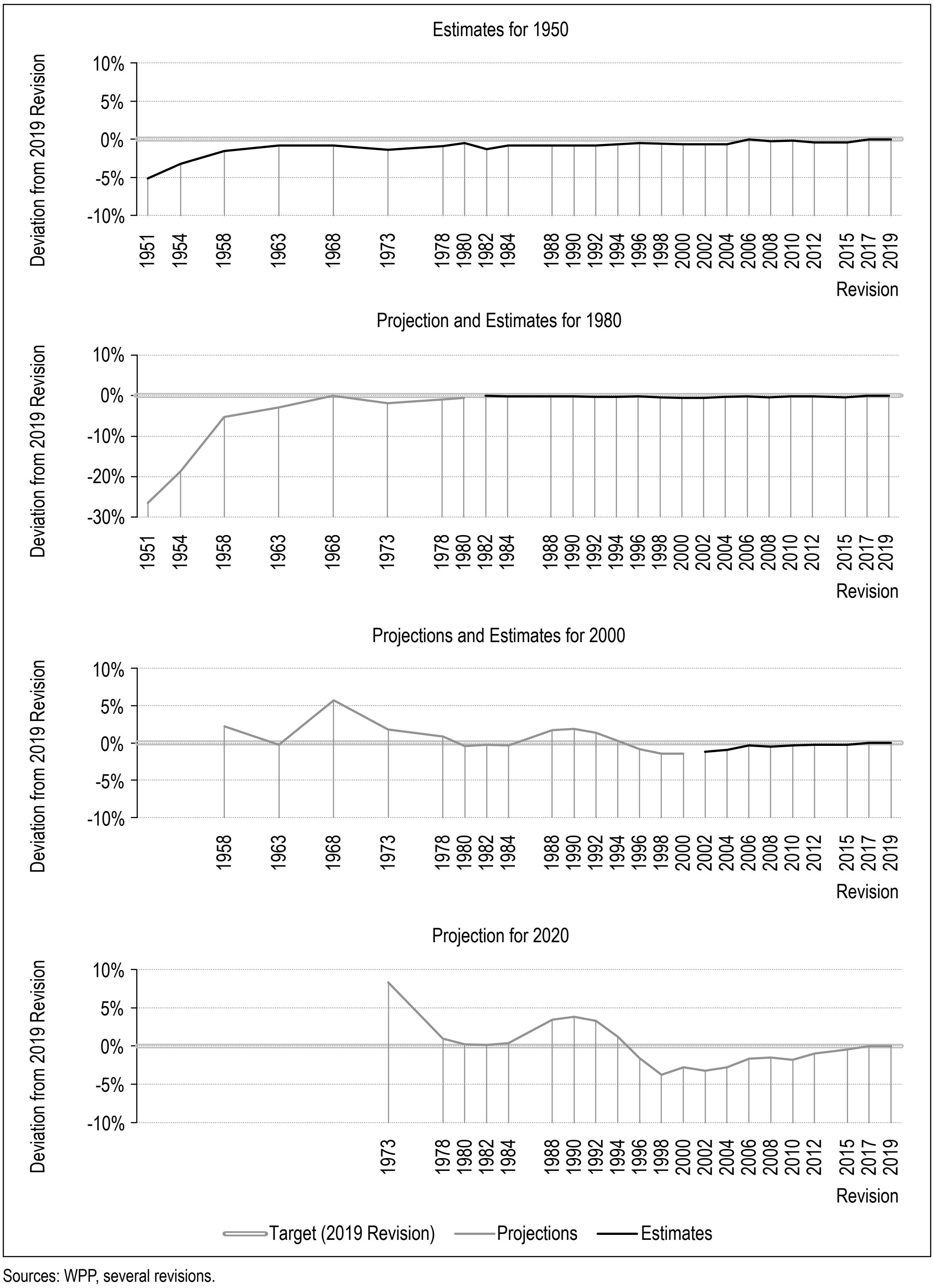
17. This necessity reflects, in part, the established usual workflow at the Population Division – countries are assigned to individual demographers in a first step, and then aggregated in a second step.

18. It is fair to assume that between 50 and 80% of the work invested in each revision is devoted to analyze, establish and revise past estimates, including the current projection's base population.

2000 and 2020 – for all past revisions and the figures published with the 2019 Revision.

For instance, the 1951 Revision estimated a population of 2.406 billion people for the year 1950, while the latest estimate for the year 1950 according to the 2019 Revision is 2.536 billion. This amounts to a 5.1% underestimation of the initial estimate compared to the current estimate. Ex-post adjustments for some countries were significantly larger (but not shown). The relative adjustments for the world population for the calendar years 1950, 1980, 2000 and 2015 are shown in Figure VI. For calendar year 1950, all

Figure VI – **Hitting the target: World population at 1950, 1980, 2000, and 2020, by revisions**



Sources: WPP, several revisions.

data are (revised) past estimates by all revisions, while for the chart showing the data for the year 2020, all data point are projections from past revisions (including the 2019 Revision). For the years 1980 and 2000, the data are either projections (for revisions prepared before that year) or estimates (for revisions prepared after that year).

It is interesting to note that all revision initially underpredicted the calendar years 1950 and 1980 before closing in on the reference numbers of the 2019 Revision. The history of approaching the 2019's reference number for the calendar years 2000 and 2020 are showing less variations but include positive and negative deviations.

Past WPPs incurred much larger errors for individual countries through missing, incorrect, or manipulated population statistics. Two examples are Bhutan, a medium sized country, and the populous African country Nigeria; both had to be corrected significantly in the past revisions (Figure VII).

Bhutan's population was completely revised in the 2006 Revision, reducing its population size dramatically. The story goes back to the beginning of the 1970s, when Bhutan joined the UN and reported a population of about 1 million inhabitants, based on a 1969 census. Because of the lack of follow-up censuses, the initial figure of about 1 million was backward projected to 1950 and forward projected assuming reasonable growth rates. It was not before the preparation of the 2006 Revision that new information from the 2005 Census became available that suggested a gross overestimation of Bhutan's past population. The initial figure of 1 million in 1970 (according to the 1973 Revision) was corrected to 297 thousand inhabitants in 1970, to less than one third. This affected the base population of the subsequent revisions; for the 2019 Revision, the figures for 2020 changed from about 2.1 million prior to the 2006 Revision to almost a quarter or 591 thousand inhabitants.

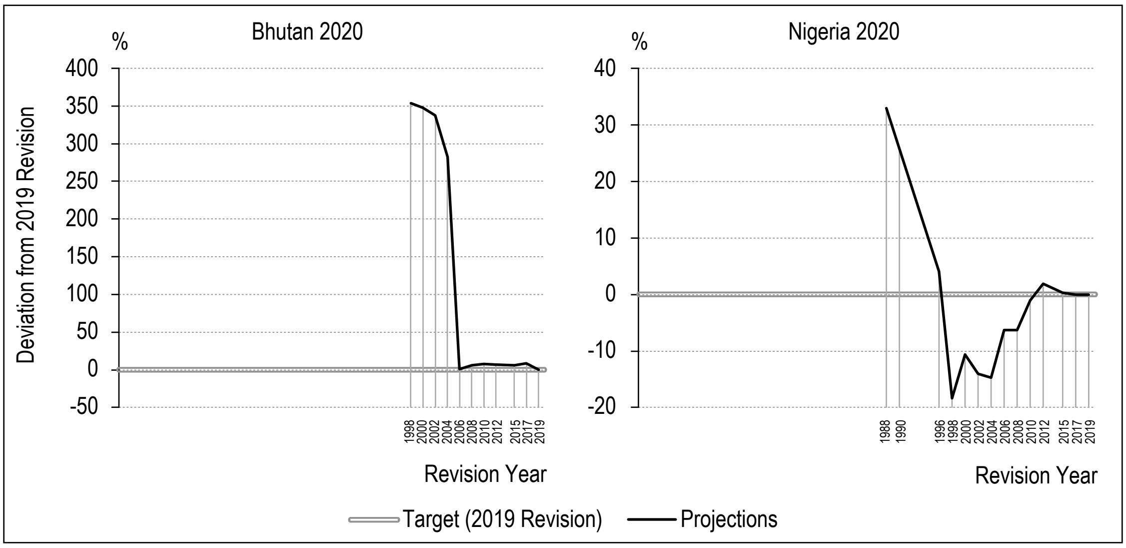
Another long-standing controversy about the "true" population figures of Nigeria is also reflected in the various revisions. All censuses of 1963, 1991 and 2006 censuses were found to need substantial adjustments for underenumeration.

The large fluctuations for the estimated population figures of Bhutan and Nigeria are exceptional, but smaller errors are common. For the world population, many of the variations are cancelling each other out. For the countries, establishing the true population estimates remains a challenge.

Projecting populations by age group and sex requires assumptions about the future course of fertility, mortality, and migration.[19] All three components must be prepared for all relevant age groups and by sex. Preparing these components of future population change was only feasible by developing and using mathematical models of trends and age patterns. The production of population projections is also, partly, informed by expert opinions, both from the outside (through workshops, etc.) and from inside the

19. *This is, of course, also true for past estimates that are themselves projections.*

Figure VII – **Hitting the target: Total populations of Bhutan and Nigeria in 2020, by revisions**



Sources: WPP, several revisions.

Population Division. The Population Division was also following technological advances, not always swiftly, by employing electronic computing devices (from mainframe computer to workstations to database server farms), plus various software, often developed in-house. Such technological progress was a significant factor in improving and expanding the scope of WPP, but it was also a tremendous challenge for staff and budgets. Change at every level and component was constant.

Indeed, it is the future population that is destined to attract the most attention every time a new revision is published. As long as the demographic transition is not complete, and the demographic momentum of many developing countries operates, continued population growth, at the world's level and for some regions, is easy to communicate, at least for now. But the prediction intervals now attached to the regular WPP projections suggest a less certain outcome than was anticipated in the past.

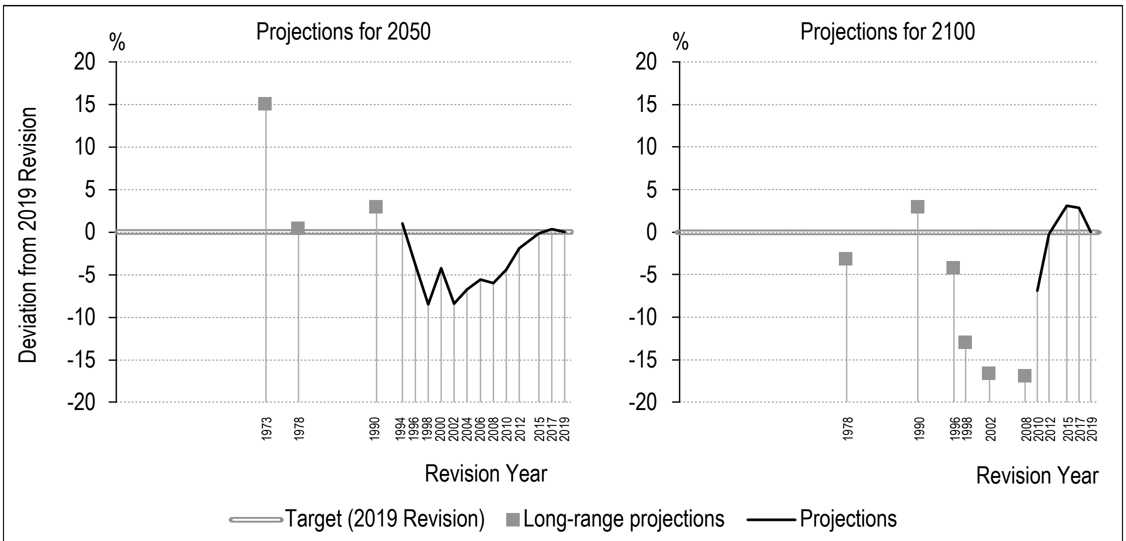
Some have argued, based on UN projections, that world population growth will continue until the end of the century (Gerland *et al.*, 2014). Others disagree (Lutz & KC, 2010; Lutz *et al.*, 2001). Some degree of projection uncertainty is apparent and justified, even between different producers and users of population projections. It is true that the United Nations Population Division maintained for a long time the replacement level of fertility – at about 2.1 children per woman – as an ultimate limit. The vision of population stabilization seemed not only a plausible, realistic, and neutral outcome. Alternative outcomes would be unsustainable population growth or continued decline. It may also be argued that population stabilization is a vision that countries in their different stages of the demographic transition could more easily accept.

How has the United Nations performed in projecting the future of global population growth? Focusing on projections up to the years 2050 and 2100, respectively, a comparison is made first with the results of past revision against the current revision of 2019. This assumes, implicitly, that the last such projection is more plausible than its predecessors, which may be doubted. But the accumulated evidence that the last revision had access to, and the methodological improvements make this assumption a plausible one.

Figure VIII depicts the total world population for the years 2050 and 2100, respectively, as produced by several past revisions. Included are early long-range projections (shown as data points), and regular projections (formatted as line) and referenced by the revision year. The x-axis is therefore not showing calendar years, but revision years. The comparison with the 2050 and 2100 projections results are shown as relative to the latest revision's projected figure as percentages.

Figure VIII shows, for the year 2050, that even relatively early projections (the long-range projections based on the 1978 Revision) came remarkably close to what the 2019 Revision established. But it also shows that an early attempt, based on the 1973 Revision long range projections, was off the mark by staggering

Figure VIII – **Approaching the future: World population in 2050 and 2100 by revisions**



Sources: WPP 1998 through 2019.

15% or 1.5 billion. The regular revisions with a projection horizon of 2050 or later were mostly underprojecting the world population in 2050, by up to 8% (1998 and 2002 Revisions). The last three revision (2015, 2017 and 2019) all produced very similar results of about 9.7 billion people in 2050.

The comparison of the various projections results up to the year 2100 showed more variation, due in part to the longer projection period. Most long-range projections[20] came out with results significantly lower that the reference from the 2019 Revision. Regular projections up to 2100, starting with the 2010 Revision, exhibited relatively small variation of less than 5%.

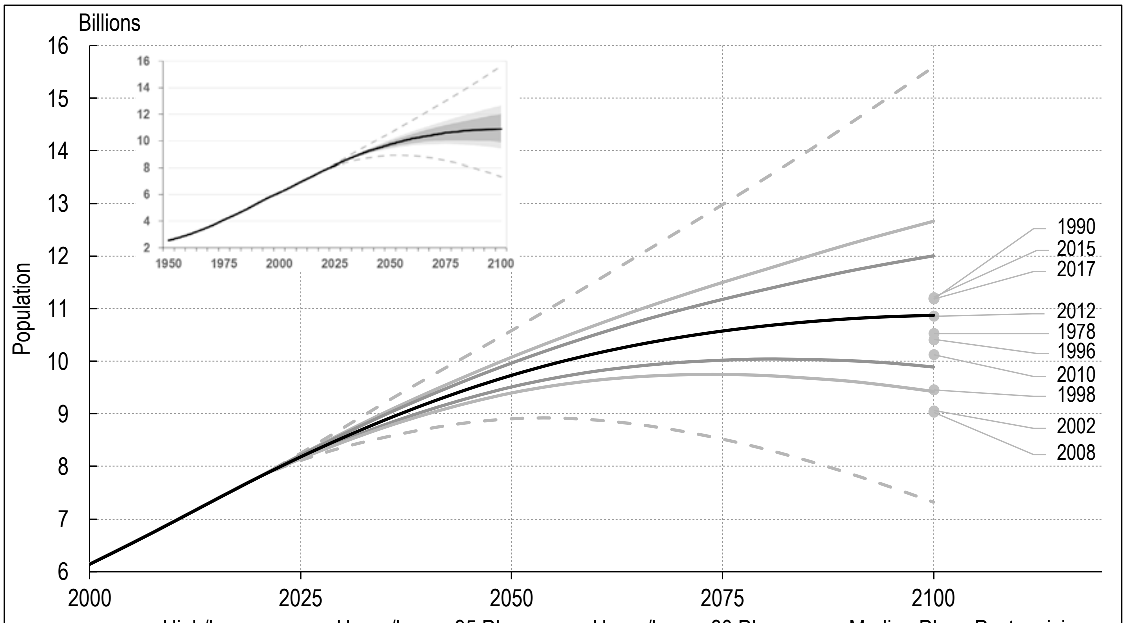
Another way to gauge how past revisions compare with the high-low variants and the prediction intervals of the 2019 Revision is shown in Figure IX. The figures for the world from 2020 to 2100 by high/low variants and prediction intervals illustrates the increasing uncertainty of that projection. The inclusion of long-range projection (1978 though the 2008 Revisions), and regular projections (2010 to 2017 Revision) shows that most of these earlier projections for the year 2100 landed with their medium variants within the 80% prediction intervals of the 2019 Revision.

Before the transition to probabilistic projections, the UN used a quite simple device to illustrate the inherent uncertainty of its projections. For most of its revisions, it defined a high and low variant that, after a short transition period after the base year, added or subtracted 0.5 children to the medium fertility variant.[21] In other words, there is a range of one child presumed to cover uncertainty. Such a uniform assumption of a fixed bound is neglecting many factors contributing to the uncertainty of future fertility. But it is easy to communicate and easy to understand. How do the past high-low ranges compare with Bayesian prediction intervals in terms of population numbers? The answer is mixed.

It appears that the 95% prediction interval and the traditional High-Low variants are similar for countries now exhibiting level of fertility between (roughly) two and three children per woman. For countries with higher fertility, the traditional high-low variants underestimate the range of possible outcomes. In contrast, for countries with below replacement fertility, the high-low variants overestimate uncertainty – the prediction intervals are much narrower. These results are plausible: low fertility countries at the end of the fertility transition are more likely to exhibit much smaller changes in fertility levels. For those countries still undergoing the transition from high to low fertility, there is more change possible. A simple comparison of the classic and the probabilistic approach is shown in Table 11.

20. Long range projection results for the years 2050 and 2100 are shown as data points to distinguish those from regular projections.
21. There was no variation in the assumed path of future mortality.

Figure IX – **Comparing the futures: World population to 2100 by revisions and prediction intervals**



Sources: WPP 1998 through 2019.

Table 11 – **Comparison of high/low population projection variants with 95% prediction intervals**

	Number of countries		% of world population	
	2050	2100	2050	2100
High and low variants are outside the 95% bounds	43	49	29	22
High or low variant is outside a 95% bound	51	71	13	16
High and low variants are within the 95% bounds	107	81	58	62
Total	201	201	100	100

Sources: WPP 2019, author's calculation.

For 43 countries in 2050 and for 49 countries in 2100, the high and low variants are indicating an uncertainty range exceeding the 95% prediction interval. For a relatively large number of countries, the classical high and low variants underestimate the 95% prediction intervals.

* *
*

What is next? Having reconstructed the evolution of the world population for the past 70 year, from 1950 to 2020 and producing, since the 2010 Revision, regular population projections to the end of this century, what could be improved, added, changed?

First what should be kept.

The reconstruction of the world's demographic history (the past estimates) from 1950 to 2020 is an asset. Now available for 70 calendar years, it provides a complete, internally consistent, and accessible data base with a host of demographic indicators for all countries, by age and sex. This database is the product of decades-long efforts to analyze, adjust, and update existing empirical sources, and close data gaps where they existed. It should be maintained and expanded. If feasible, an extension back to the beginning of the 20th century would certainly be much welcome by historians, epidemiologist, economist, and many more. The Population Division could build upon its expertise and existing, if still fragmented, data collections.

There is a natural tension between official statistics and estimates produced by non-state actors. Official statistics, produced by government authorities, are a political statement about a country's situation. Statistical estimates by international organizations, for instance, are produced independently, using additional sources of data and occasionally alternative methodology and methods. Tensions between official statistics and independent estimates arise when official statistics are incomplete[22] or use concepts and definitions that are not internationally comparable. Therefore, the Population Division often revises official statistics in terms of concepts, comparability, and consistency. Since past estimates are produced by cohort-component projections starting at 1950, they guarantee internal consistency across the time, age, and sex dimensions. As was shown, that internal consistency – a remarkable achievement – is also constantly under review. Thus, estimates produced by the Population Division are adding utility to the international statistical system, but do not replace official statistics.

Already in 1947, Trygve Lie, first Secretary-General of the United Nations, expressed the need for consistent and comparable population estimates for the United Nations System: "On one point we believe the central organization has a special obligation. Confusion could arise if a variety of slightly different population estimates were used by the various organizations. I suggest, therefore, that, in so far as possible, the United Nations should be called upon to provide the current estimates of population used throughout the various organizations […] In the most general terms, it will be the special duty of the Statistical Office to assure the flow from and to governments of basic data in the field of demography as in other fields. The Office also has special obligations in matters of statistical methods and standards. The Population Division, on the other hand, has the major obligation for investigation and analysis. Between these two fields of competence there is a considerable area; but we have decided not to delimit the boundaries more precisely at present." (United Nations, 1995, p. 870).

Ultimately, official statistics and the estimates produced by the Population Division should not be seen as competing, but as presenting a picture of the world, but with different objectives. The independence of WPP, especially its estimates, from official statistics should be maintained.

22. Areas of disagreement between some official statistic and the estimates prepared by the United Nations Population Division are the treatment of census undercounts, especially children and, sometimes, women (measured or inferred) and the retroactive correction of past estimates after a census.

What could be added or improved?

Migration has always been the most problematic element in international population projections. Recent methodological advancements (Abel, 2013, 2016; Abel & Sander, 2014; Azose & Raftery, 2019; Buettner & Muenz, 2018a, 2018b), plus the results of the decadelong efforts to collect, review and adjust migrant stock data (from censuses) and migration flows (from selected countries), have made it now possible to include migration in a much more transparent and policy-relevant fashion: as flows of people between countries.[23] This would be, no doubt, a challenging and resource-demanding project, and it is probably a longer-term enterprise. The Population Division could follow their own example by building the evidence patiently and consistently, and in cooperation with other agencies, organizations, and the academic community. Implementing migration flows into the WPP would be an important improvement.

Demographic projections are necessarily uncertain. While this was accepted from the beginning of the WPP exercise, there were different attempts to account for that uncertainty. In most of WPP's history, some measure of uncertainty has been constructed by calculating high and low variants around a central or medium variant, almost exclusively for fertility levels. Such a naïve approach reflected the lack of detailed data (for single calendar year, for instance) and the weak computing power of those times. The recent shift to a complex and sophisticated probabilistic projections model[24] based on hierarchical Bayesian models is a significant progress in this regard. It has also made interpretation and communication of results much more complex.

Keyfitz' warning against the misuse of projections variants rings also true for probabilistic projections: "If […], as more commonly happens, the user looks at the results and takes whichever of the three projections [low, medium and high] seems to him most likely, then the demographer has done nothing for him at all – the user who is required to choose on the basis of which of the results looks best might as well choose among a set of random numbers." (Keyfitz, 1981, p. 591).

But how to communicate uncertainty? Is one to favor the median results or the confidence margins instead? The 2019 Revision tried this: "Although the most likely scenario is that the world's population will continue to grow throughout the present century, there is an estimated 27 per cent probability that it could stabilize or even begin to shrink sometime before 2100".

In order to make the results of probabilistic projections more accessible, demographers have suggested "[…] in order to achieve a paradigm shift in practical applications of probabilistic population forecasts, the focus should not be on methods, but rather on possible impacts and consequences of decisions." (Bijak *et al.*, 2015, p. 542). The issue of handling and communicating uncertainty remains work in progress and must be developed further.

The outputs of current WPPs are impressive: Volumes with Key Findings, Comprehensive Tables, Demographic Profiles, Methodology; Data Booklets, Wallcharts, related technical papers and population facts; a complete online presence of the results, plus an interactive database, online documentation of data sources, interactive charts, thematic maps and data files in different formats for the occasional and the power user. Quite impressive but requiring a huge amount of resources.[25]

The issue of an optimal time schedule for publishing new/revised estimates and projections have been discussed already in the past (United Nations, 1984, p. 4). It may be worthwhile to reopen the discussion about how to react to new evidence and new or improved methodology while optimizing the amount and depth of results in its many forms. A careful reader may sometimes recognize some paragraphs in a new revision that are copied *verbatim* from a previous one. It seems recommendable to restrict some updates to electronic media and update the printed copies at longer intervals to reduce the burden of the demographers to produce lengthy documents that contain numerous repetitions.

Apart from the volume of publishing results, the frequency of updates appears to be a challenge for some users and might even have detrimental effects on data collection systems (Boerma *et al.*, 2018)., and the re-estimation of estimates is also not always welcome (Rigby *et al.*, 2019). It will remain a challenge for the demographers at the United Nations to find a balance between completeness, timeliness, and feasibility.

In conclusion, looking back at an impressive history of 70 years providing an authoritative account of the world's demography since 1950, and increasingly informative projections,

23. *A first attempt to include flows into international population projections was made by Lutz* et al. *(2014).*
24. *The inclusion of explicit probabilistic measures of uncertainty into population projections has been long suggested by many demographers (Ahlburg et al., 1998; Keilman et al., 2002; Lutz & KC, 2010).*
25. *The actual manpower producing the Population Division's estimates and projections is surprisingly small.*

the Population Division's World Population Prospects remain an important and valuable project that will evolve and improve in the future. For the world is not a Panglossian paradise but needs work to improve. *Il faut cultiver notre jardin.* ☐

Link to the Online Appendices: https://insee.fr/en/statistiques/fichier/4997857/ES-520-521_Buettner_Online_Appendices.pdf

BIBLIOGRAPHY

Abel, G. J. (2013). Estimating global migration flow tables using place of birth data. *Demographic Research*, 28(March), 505–546. https://doi.org/10.4054/DemRes.2013.28.18

Abel, G. J. (2018). Estimates of Global Bilateral Migration Flows by Gender between 1960 and 2015. *International Migration Review*, 52(3), 809–852. https://doi.org/10.1111/imre.12327

Abel, G. J. & Sander, N. (2014). Quantifying Global International Migration Flows. *Science*, 343(6178), 1520–1522. https://doi.org/10.1126/science.1248676

AbouZahr, C., Cleland, J., Coullare, F., Macfarlane, S. B., Notzon, F. C., Setel, P., … & Zhang, S. (2007). The way forward. *The Lancet*, 370(9601), 1791–1799. https://doi.org/10.1016/S0140-6736(07)61310-5

AbouZahr, C., de Savigny, D., Mikkelsen, L., Setel, P. W., Lozano, R. & Lopez, A. D. (2015). Towards universal civil registration and vital statistics systems: the time is now. *The Lancet*, 6736(15), 1407–1418. https://doi.org/10.1016/S0140-6736(15)60170-2

Ahlburg, D. A., Lutz, W. & Vaupel, J. W. (1998). Ways to Improve Population Forecasting: What Should Be Done Differently in the Future? *In:* D. A. Ahlburg, W. Lutz & J. W. Vaupel (Eds.), *Frontiers of population forecasting*, pp. 191–199. http://www.jstor.org/stable/2808056

Azose, J. J. & Raftery, A. (2019). Estimation of emigration, return migration, and transit migration between all pairs of countries. *Proceedings of the National Academy of Sciences*, 116(1), 116–122. https://doi.org/10.1073/pnas.1722334116

Bijak, J., Alberts, I., Alho, J. M., Bryant, J., Buettner, T., Falkingham, J., … & Smith, P. W. F. (2015). Probabilistic Population Forecasts for Informed Decision Making. Letter to the Editor. *Journal of Official Statistics*, 31(4), 537–544. https://doi.org/10.1515/jos-2015-0033

Biraben, J.-N. (1979). Essai sur l'évolution du nombre des hommes. *Population (French Edition)*, 34(1), 13–25. https://doi.org/10.2307/1531855

Biraben, J.-N. (2006). The History of the Human Population from the First Beginnings to the Present Day. *In: Demography: Analysis and synthesis. A treatise in demography*, Vol. III, pp. 5–17.

Boerma, J. T., Victora, C. & Abouzahr, C. (2018). Monitoring country progress and achievements by making global predictions: is the tail wagging the dog? *The Lancet*, 392(10147), 607–609. https://doi.org/10.1016/S0140-6736(18)30586-5

Buettner, T. & Muenz, R. (2018a). International Migration Projections: Methodology Brief. KNOMAD *Working Paper* N° 30 – Annex. Washington, D.C.

Buettner, T. & Muenz, R. (2018b). Modeling Alternative Projections of International Migration. KNOMAD *Working Paper* N° 30. Washington, DC. https://www.knomad.org/publication/modeling-alternative-projections-international-migration

Caswell, H. (2001). *Matrix Population Models: Construction, Analysis, and Interpretation* (Second Ed.). Sunderland, Mass.: Sinauer Associates.

Davis, K. (1945). The World Demographic Transition. *The Annals of the American Academy of Political and Social Science*, 237(1), 1–11. https://doi.org/10.1177/000271624523700102

Durand, J. D. (1974). Historical Estimates of World Population: An Evaluation, (10). https://repository.upenn.edu/psc_penn_papers/9/

Gerland, P., Raftery, A., Sevcikova, H., Li, N., Gu, D., Spoorenberg, T., … & Wilmoth, J. R. (2014). World population stabilization unlikely this century. *Science*, 346(6206), 234–237. https://doi.org/10.1126/science.1257469

Keilman, N., Quang Pham, D. & Hetland, A. (2002). Why population forecasts should be probabilistic – illustrated by the case of Norway. *Demographic Research*, 6, 409–154. https://doi.org/10.4054/DemRes.2002.6.15

Keyfitz, N. (1972). On Future Population. *Journal of the American Statistical Association*, 67(338), 347–363. https://doi.org/10.1080/01621459.1972.10482386

Keyfitz, N. (1981). The Limits of Population Forecasting. *Population and Development Review*, 7(4), 579–593. https://doi.org/10.2307/1972799

Landry, A. (2020). *La révolution démographique. Études et essais sur les problèmes de la population*. Paris: Éditions de l'Ined - Collection Classiques de l'économie et de la population.

Lutz, W., Butz, W. P. & KC, S. (Eds.) (2014). *World population and human capital in the twenty-first century*. Oxford: Oxford University Press.

Lutz, W. & KC, S. (2010). Dimensions of global population projections: what do we know about future population trends and structures? *Philosophical Transactions of the Royal Society B*, 365, 2779–2791. https://doi.org/10.1098/rstb.2010.0133

Lutz, W., Sanderson, W. C. & Scherbov, S. (2001). The end of world population growth. *Nature*, *412*, 543–545. https://doi.org/10.1038/35087589

Mikkelsen, L., Phillips, D. E., AbouZahr, C., Setel, P. W., de Savigny, D., Lozano, R. & Lopez, A. D. (2015). A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. *The Lancet*, 6736(15), 1395–1406. https://doi.org/10.1016/S0140-6736(15)60171-4

Notestein, F. W. (1945). Population – The Long View. *In:* T. W. Schultz (Ed.), *Food for the World*, pp. 36–57. Chicago: University of Chicago Press.

Population Commission. (1947). *Report to the Economic and Social Council on the First Session of the Commission Held at Lake Success*. New York, 6 to 19 February 1947.

Rigby, M., Deshpande, S. & Blair, M. (2019). Credibility in published data sources. *The Lancet*, 393(10168), 225–226. https://doi.org/10.1016/S0140-6736(18)32844-7

Setel, P. W., Macfarlane, S. B., Szreter, S., Mikkelsen, L., Jha, P., Stout, S. & AbouZahr, C. (2007). A scandal of invisibility: making everyone count by counting everyone. *The Lancet*, 370(9598), 1569–1577. https://doi.org/10.1016/S0140-6736(07)61307-5

United Nations (1948). *Yearbook of the United Nations 1947-48*. Lake Success, New York: UN. https://doi.org/10.18356/72ee6d65-en

United Nations (1949). *World population trends, 1920-1947*. New York: Department of Economic and Social Affairs, Population Division.

United Nations (1951). *The Past and Future Growth of World Population - A Long Range View. Population Bulletin, N°1*. New York: Department of Economic and Social Affairs, Population Division.

United Nations (1984). Population Projections: Methodology of the United Nations. *Population Studies* No. 83. New York: Department of Economic and Social Affairs.

United Nations (1995). Trygve Lie on Population in the United Nations' Agenda. *Population and Development Review*, 21(4), 867–873. https://doi.org/10.2307/2137778

United Nations (2019a). How certain are the United Nations global population projections? *Population Facts*. New York: Department of Economic and Social Affairs, Population Division. https://doi.org/10.1073/pnas.1713628115

United Nations (2019b). *World Population Prospects 2019: Methodology of the United Nations population estimates and projections* (ST/ESA/SER.A/425).

Vallin, J. & Caselli, G. (2006). The United Nations' World Population Prospects. *In: Demography: Analysis and synthesis. A treatise in demography*, Vol. III, pp. 197–233.

Bayesian Probabilistic Population Projections for France

Vianney Costemalle*

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%.

JEL Classification: C11, C53, J11, J13, F22
Keywords: probabilistic projections, Bayesian inference, time series, population, mortality, fertility, migration

**Insee at the time of writing (vianney.costemalle@sante.gouv.fr)*
Acknowledgments – The author would like to thank Julian Arbel, Junni Zhang, John Bryant, Marie Reynaud, Isabelle Robert-Bobée, Nathalie Blanpain, Guillemette Buisson, Vanessa Bellami, Catherine Beaumel as well as two anonymous referees.

Received November 2017, accepted July 2018.
Translated from: “Projections probabilistes bayésiennes de population pour la France”
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

Population projections are performed regularly by statistics institutes around the world, as well as certain international organisations such as the United Nations (UN), which has published *World Population Prospects* (UN, 2017) every two or three years since 1951. Population projections offer many benefits and they have numerous users. They are primarily used to predict the possible future population of a region, a country or the entire world under certain assumptions, in terms of both number of inhabitants and structure. In the short to medium term, these projections form the basis for economic and social planning, such as pension funding (COR, 2017) or the construction of public infrastructure. They also form an essential part of certain other exercises, such as economic, climate or environmental projections.

In the case of France, the most recent official projections date from 2016 (Blanpain & Buisson, 2016a ; 2016b) and provide an indication of what the population will be in 2070 if past trends continue, with different variants on these assumptions (see Blanpain in this issue). Details of the projections by region, and in particular those for metropolitan France, are only available for the period from 2013 to 2050. This article aims to explore a new method for projecting the population of France: probabilistic projections. The proposed approach is said to be probabilistic since it allows the uncertainty surrounding future population levels to be quantified. This is where it differs from the traditional approach, which is a set of deterministic projections based on different scenarios. The fundamental difference between these two approaches is not so much the results themselves, but the way in which they are interpreted and used.

Probabilistic projections are based on statistical models, the majority of which are parametric. The uncertainty surrounding some elements making up the population can be captured by error terms, as is the case with time series, but it can also come from Bayesian inference of the model's parameters. The aim is to quantify the level of uncertainty surrounding the future population. This can be achieved using the stochastic approach, the Bayesian approach or even a combination of the two. In this article, we use stochastic models with Bayesian inference of the parameters.

In a letter to the editor of the *Journal of Official Statistics*, a group of demographers and academics from various countries highlighted the contributions and challenges of probabilistic projections in demography and called for more research and practice in this area by statistics institutes (Bijak *et al.*, 2015). They highlighted the fact that probabilistic projections have already been developed and used successfully in other disciplines, such as meteorology, climatology and even aviation. Bayesian statistics are also taking their time in breaking through into the field of demography. Although Bayes' theorem was established more than 250 years ago, it is only recently, with the appearance of MCMC (Markov Chains Monte-Carlo) algorithms in the 1980s and the explosive increase in computer processing power, that Bayesian inference has been used (Bijak & Bryant, 2016).

Some statistical institutes have already adopted the approach aimed at producing probabilistic population projections for their official statistics. This is the case in the Netherlands and New Zealand in particular. The Netherlands started producing probabilistic projections based on stochastic methods in 1998. New Zealand has also been reporting probabilistic population projection results since 2012 (MacPherson, 2016; Dunstan & Ball, 2016). The UN, which develops projections for all countries, eventually switched from a deterministic to a probabilistic method in 2014 (Costemalle, 2015). Furthermore, some elements of its projections are based on Bayesian inference.

The overwhelming majority of population projections are based on the component method, which consists of producing separate projections for the three key components of population dynamics, namely fertility, mortality and migration. The population at a given time is broken down into sex and age categories and is equal to the population during the previous period plus births and immigrants and minus deaths and emigrants. In this way, it is possible to chart the development of the population and its structure by sex and age category from one period to the next. In order to achieve this, the number of births by sex must be determined for each period, along with the number of deaths and the net migration by sex and age group. As regards births and deaths, the most common methods are based on projected fertility and mortality rates. However, probabilistic population projections remain an active area of research: there is no single method; on the contrary, there are almost as many approaches as there are types of data and they differ from one country to the next.

In the first part of this article, we will highlight the key differences between deterministic and probabilistic projections before going on to describe some of the different approaches that have been developed in demography with respect to probabilistic population projections. The second part is dedicated to describing French mortality, fertility and net migration data, and the third part looks at the presentation and validation of the models used for each of the three components. We finish by presenting the results of the probabilistic projections obtained in this manner for France, before going on to discuss the assumptions within the models.

1. Deterministic and Probabilistic Projections and Developments in Demography

Predicting the future is a difficult task and has given rise to the development of many different methods over the centuries. The most recent and sophisticated methods are based on mathematical models that attempt to detect certain patterns or invariants in the data and to extrapolate the trends observed, while also respecting certain constraints that may be imposed. Both deterministic and probabilistic projections require a certain degree of modelling of observed data and differ only in the nature of the forecasts made.

1.1. Deterministic and Probabilistic Approaches: Different Ways of Addressing the Future

In the first instance, what we are looking to project depends, from a deterministic standpoint, on certain parameters. The selection of these parameters represents a hypothesis that is also referred to here as a scenario. A scenario is then given detailing the way in which these parameters are considered most likely to develop on the basis of accumulated knowledge, expert opinions and intuition. A given scenario corresponds to one single possible projection, and the relationship between the two is deterministic. In cases where the scenario plays out as expected, the projection will be certain. Deterministic projections answer the question: “What would happen in the future if such a scenario were to occur?”. Extreme scenarios can therefore be created to see how the future would pan out if they were to come true. Deterministic projections are thus a formidable tool when it comes to exploring the future on the basis of predefined scenarios. Any uncertainty in the projection then relies on the scenario coming true. Possible scenarios are formulated, but it is impossible to know how likely they are to occur. It could even be argued that the probability of them coming true is zero (if the values are continuous) or very low (if the values are discrete). The degree of probability is estimated intuitively and is reflected in the terms used to describe these scenarios: demographers refer to the “central” scenario, which is the scenario considered the most plausible based on current knowledge, and “extreme” scenarios.

Conversely, probabilistic projections are based on models that attempt to take account of the uncertainty stemming from a lack of knowledge of certain aspects of the projections. These models are based on assumptions made on the basis of expert judgement and intuition. The underlying assumptions on which models for probabilistic projections are based are the equivalent of the scenarios used for deterministic projections. The advantage of probabilistic projections is that they make it possible to quantify the uncertainty based on past developments and to extrapolate it into the future to provide confidence intervals for the projections. The interpretation and use of probabilistic projections therefore differs from that of deterministic projections.

By way of an example, weather forecasts have long been making use of probabilistic projections: we are not only told whether or not it will rain the next day, but also the probability that rain will fall (Raftery, 2014). Since future events are inherently uncertain, indicating the probability of their occurrence in view of current knowledge provides more information than a deterministic projection based on a scenario. In economics in particular, time series are used as a means of producing probabilistic projections: in the case of a simple random sampling method, for example, we know that the variance increases with the square root of time.

By adding error terms to the models, it is therefore possible to create stochastic probabilistic projections. Another method for quantifying uncertainty is to use the Bayesian paradigm. Under this method, the model parameters are viewed as random variables, in the same way as error terms in stochastic models. Bayesian inference then involves estimating the *a posteriori* distribution of these parameters, i.e. after the data have been observed. This distribution gives possible values for the parameters, together with their degree of probability. It differs from the *a priori* distribution, which is the distribution given by the modeller and which is intended to reflect the knowledge of the problem before any data has been observed.

1.2. Probabilistic Projections in Demography: A Wide Variety of Models in Practice

Population projection techniques can be divided into three categories (Booth, 2006). The first group includes methods based on the extrapolation of trends, which seek to extend the trends identified in the past, in most cases in a linear fashion. They are based solely on past data and do not attempt to explain the mechanisms underlying the developments. They often prove to be effective. The second set of methods used for population projections involves establishing long-term trends. These methods are based on the expectation that the future will unfold in a certain way. This may be backed up by expert opinions, which assess what could be expected to happen in the future on the basis of current knowledge, or on people's intentions, such as those measured by fertility intention surveys (Régnier-Loilier & Vignoli, 2011). Finally, the last category of projections is made up of the structural models, which attempt to explain the mecanisms of population changes using exogeneous variables. These exogeneous variables must then be projected in accordance with one of the three projection categories. The approaches often combine several of these techniques and the techniques used differ according to the components (mortality, fertility and migration) that are to be projected.

A classic method of projecting mortality was developed by Lee & Carter (1992) and consists in decomposing the change in the logarithm of mortality rates into an age effect and a time effect, specific to each age. The time effect is then considered as a time series for which the parameters are estimated. By calculating or simulating the future values of this time effect on the basis of the models used a very large number of times, it is possible to obtain a probabilistic projection. The basic idea of this approach is to capture the regular changes in the data and to extrapolate these regularities. The Lee-Carter method has since been used very frequently to project mortality, as well as to project fertility and migration. Wiśniowski *et al.* (2015) put forward a more extended version of this, adding a generation effect, which can be applied to all three components of population change. In addition, these authors have proposed that these projections be carried out in an entirely Bayesian framework. The Lee-Carter model has also been generalised by Hyndman & Ullah (2007), who break down the logarithm for mortality rates or fertility rates into key components before extending the coefficients of each of those components using time series. Furthermore, Hyndman & Booth (2006) suggest performing a Box and Jenkins transformation on the rates studied with a view to generalising the log transformation. This approach is entirely stochastic.

The whole point of probabilistic projections is to allow the degree of probability of future projections to be quantified. In 2001, Lutz *et al.* (2001) announced that the world population is likely to stop growing by the end of the century. More specifically, their stochastic models and calculations predict that there is an 85% probability that the world population will begin to decline by the end of the century. The UN, which regularly publishes population projections, began using a probabilistic and Bayesian method in 2014. The results give a different view of the development of the population in the long term. In fact, they show that the world population is unlikely to have stopped growing by 2100 (Gerland *et al.*, 2014). The methodology used differs from that applied by Lutz *et al.* (2001): the aggregated values, which are life expectancy at birth and the total fertility rate (TFR), are projected directly in a first step. These indicators are then decomposed in sex-specific and age-specific mortality rates and age-specific fertility rates. In order to project life expectancy, the amount by which life expectancy increases every five years is modelled by a double logit function on the basis of actual life expectancy and a large number of parameters. These parameters are estimated by Bayesian inference, which leads to an *a posteriori* distribution of increases in life expectancy and therefore an *a posteriori* distribution of life expectancy itself by 2100 (Raftery *et al.*, 2013). This is an example of a probabilistic projection that does not use stochastic terms, but is instead based solely on parametric modelling and Bayesian inference. For its part, the TFR indicator is modelled according to a three-phase process of development: a phase of high fertility rates, a phase of rapid fertility decline to below the generation replacement level, and a phase of stagnation of the fertility rate with a long-term convergence towards a level of 2.1 children per woman (Alkema *et al.*, 2010).

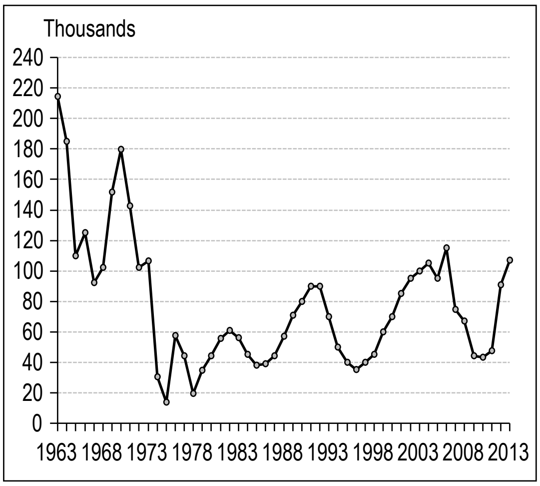
It therefore appears that there are numerous models available to project each of the three components. Working on the assumption that no single model can capture the full range of possible assumptions about mortality trends, especially when these assumptions are not consistent with one another, Kontis *et al.* (2017) made use of 21 different probabilistic projection models, the results of which were then weighted in accordance with the performance of each of the models, in order to ultimately obtain a single probability distribution for the desired indicators.

2. Data for France

To ensure that we have long series, we will restrict ourselves to the area of metropolitan France. We therefore have, for the years 1962 to 2013, the total population on 1 January of each year, the annual net migration, the number of deaths and the number of births by the age of the mother, all detailed by sex and age.[1] We have selected the same projection horizon as that used for the most recent official projections for France (Blanpain & Buisson, 2016b). The aim is therefore to project the 2014 population to 2070. Between 1962 and 1998, the data are not broken down by age beyond 100 years. From 1999 onwards they are broken down in detail up to 110 years. We then chose to retain the one-year age categories since the data are available, and we created a higher age category representing people 100 years of age and older. In the remainder of this section, we will describe the net migration, mortality and fertility data, highlighting invariants, trends and irregularities.

Net migration is the number of people in a given year who come to live in France from outside of metropolitan France, regardless of their nationality, minus the number of people living in metropolitan France who move abroad. It is undoubtedly the most difficult component to measure, because, although the number of people entering the country can be estimated using the population census (Brutel, 2014), we do not know how many have left. Net migration can therefore be calculated as the difference between the changes in the population and the natural balance. Unlike many other European countries, France does not have a population register and must therefore rely on the population census to estimate migration flows. As the census was only conducted once every 7-8 years or so until 1999, it was not possible to directly calculate the change in the population from one year to the next. In 1962, net migration was exceptionally high as a result of approximately 860,000 French nationals returning from Algeria; from 1963 onwards, net migration has been consistently positive, but the numbers have been much lower: it averaged 64,000 over the period from 1963 to 2013. Net migration appears to have remained stable on average from the 1990s onwards, although there have been some large fluctuations (Figure I), largely due to the various policies pursued, but also as a result of the economic and international context. For the period from 1990 to 2013, net migration was, on average, 72,000 and 79,000 for the last ten years available (2004-2013).

Figure I – **Changes in net migration between 1963 and 2013**



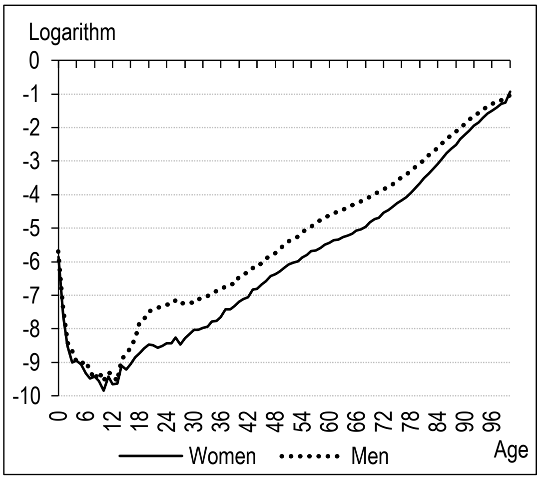
Sources and coverage: Insee, population estimates and civil registry statistics; Metropolitan France.

In order to describe mortality, the number of deaths must be related to the corresponding population at risk. This population is counted in person-years and takes account of the total time spent by all persons residing in France. It is approximately equal to the population present on 1 January, plus half of the net migration. By relating the number of deaths to this population, we then obtain the mortality rates, which can be broken down by sex, age and year. Mortality rates grow quasi-exponentially from the age of 25 upwards (Figure II). Before the age of 25, the profile is different due to infant mortality, which is higher for newborns. Mortality rates decline from birth until around 10 years of age, before rising steadily. At around the age of 18, the mortality of men becomes significantly higher than that of women, and the gap remains present throughout life, with a greater or lesser magnitude depending on age.

The logarithm of mortality rates, for a fixed age and sex, decreases in an almost linear manner over time (Figure III). This is especially true for older age groups, but does not seem to be quite the case for younger age groups. For example, the logarithm of the mortality rate at 10 years of age decreases faster and faster. Conversely, at age 30, the logarithm of the mortality rate slows its decline until it stagnates for males from the early 1980s to the mid-1990s, at which point

1. 2013 is the latest year for which all these data were final when the projections presented in this article were carried out, in 2017. In particular, the figure for net migration was not yet available for 2014. We have not used the provisional data, then available up until 2016, but which are revised from one year to the next before being final and therefore of a different nature from the final data.

Figure II – **Logarithm of mortality rates in 2013 by sex and age**



Sources and coverage: Insee, population estimates and civil registry statistics; Metropolitan France.

mortality declines sharply for that age group and has continued to decline steadily and in an apparently linear fashion since. This stagnation in mortality among young adults in the 1980s and 1990s, when the general trend was towards a steady decrease in mortality, is linked to the AIDS epidemic, which reached France in the early 1980s. In general, as mortality rates are steadily declining, life expectancy at birth is increasing each year, and more rapidly for men than for women (Blanpain, 2016), although life expectancy sometimes decreases from one year to the next, as was the case in 2015 for cyclical reasons (Bellamy & Beaumel, 2016).

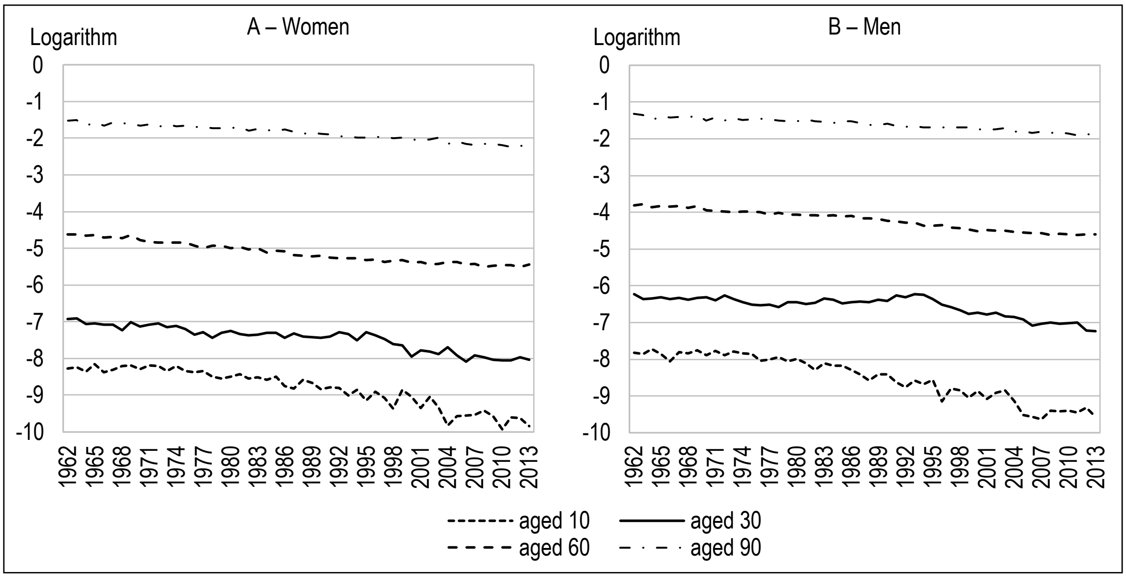
Since the early 1970s, the TFR[2] has declined sharply from 2.9 children per woman in 1964 to 1.8 children per woman in 1976 (Figure IV). It has since stabilised at an average of around 1.85 children per woman. Nevertheless, an upward trend has been observed in the TFR since the mid-1990s.

The fertility rate at a given age is defined as the ratio of the number of babies born to mothers of that age to the number of women of the same age in the year in question. This number corresponds to the number of women on 1 January of the year plus half of the corresponding net migration and minus half of the deaths recorded for this population. The profile of age-specific fertility rates follows a bell curve: the probability of having a child in a given year increases with age from 15 years until it peaks, after which it declines continuously, reaching zero or close to zero around 50 years.

Over time, this age distribution tends to shift to the right: the age at which peak fertility is reached increases (Figure V). In 1970, the fertility rate was at its highest at 24 years of age, while in 2013, the peak was reached at the age of 30. The maximum level of fertility reached during the year has barely changed since the mid-1970s: it fluctuates around 0.15. As the fertility peak moves to the right, the distribution of age-specific rates becomes increasingly symmetrical, as evidenced by the measure of skewness, which is rapidly decreasing towards 0 (Figure VI).

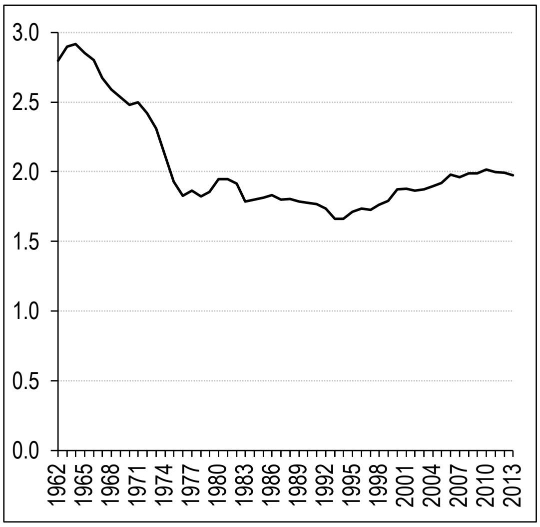
2. The total fertility rate (TFR) is calculated as the sum of the age-specific fertility rates. It corresponds to the average number of children that a woman would have during her lifetime if the probability of giving birth at a given age corresponded to the fertility rate at that age.

Figure III – **Changes in the logarithms of mortality rates from 1962 to 2013 for different ages**



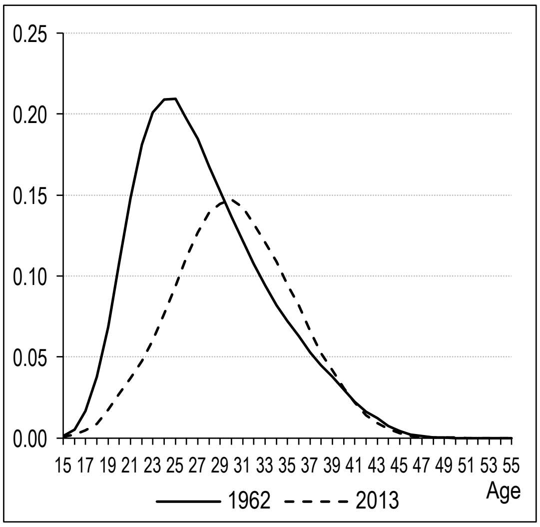
Sources and coverage: Insee, population estimates and civil registry statistics; Metropolitan France.

Figure IV – **Changes in the total fertility rate from 1962 to 2013**



Sources and coverage: Insee, population estimates and civil registry statistics; Metropolitan France.

Figure V – **Age-specific fertility rate in 1962 and in 2013**

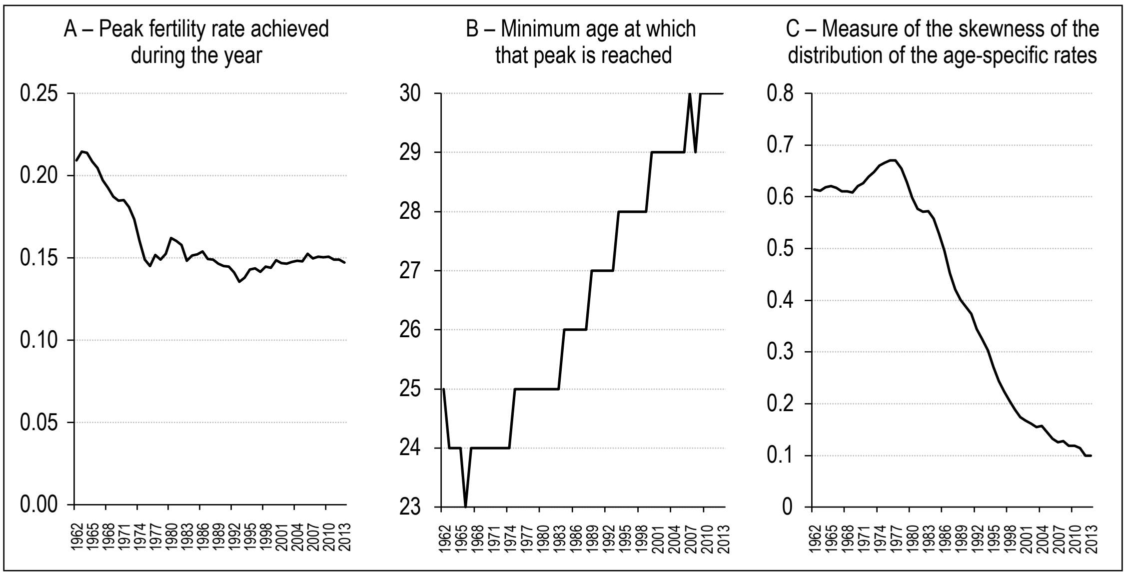


Sources and coverage: Insee, population estimates and civil registry statistics; Metropolitan France.

Unlike mortality rates, changes in fertility rates do not occur in a regular manner over time. For example, the fertility rate at 30 years of age fell between the early 1960s and the mid-1970s; however, it has been increasing since then, albeit with a slowdown from the 2000s onwards. The fertility rate at 20 years of age had been declining since the 1970s, but in the late 1990s rebounded slightly for a few years before declining again, but at a much slower pace than in previous decades. The changes are neither monotonous nor linear, which highlights how difficult it is to extend these curves into the future.

To summarise, net migration in metropolitan France appears stable over the long term, but with significant fluctuations that seem difficult to predict. Mortality has been moving in the same direction for several decades, with an almost linear decrease in the logarithm of mortality rates at all ages and a narrowing of the gap in life expectancy between women and men. Recent fertility trends are more complex to identify, but the evidence suggests that the TFR has stabilised at an average level of just under 2 children per woman and that the distribution of age-specific fertility is changing continuously with a shift in

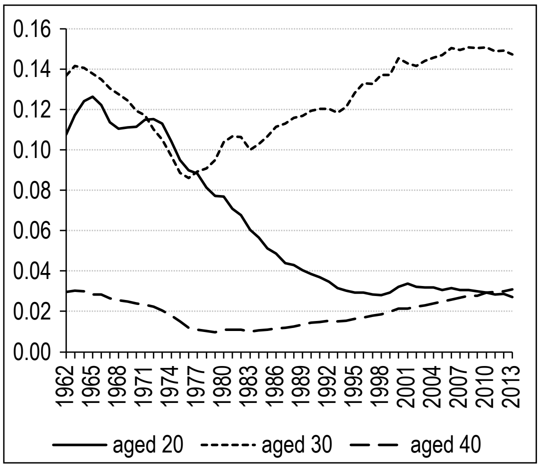
Figure VI – **Changes in peak fertility, age at which peak fertility is reached and skewness of the age distribution of fertility rates between 1962 and 2013**



Sources and coverage: Insee, population estimates and civil registry statistics; Metropolitan France.

peak fertility to higher ages and an increasingly symmetrical distribution (Figure VII). In the next section, we propose a model for each of the three components of population change, taking account of these observations and drawing upon models that have already been developed internationally and which we will describe briefly in the third section.

Figure VII – **Changes in fertility rates at different ages between 1962 and 2013**



Sources and coverage: Insee, population estimates and civil registry statistics; Metropolitan France.

3. Methods and Models

In the remainder of this article, we will use the following notations:

P(a,n,s): the number of people on 1 January of year *n*, of sex *s* born in year *n-a*;

D(a,n,s): the number of deaths in year *n*, of people of sex *s* and born in year *n-a*;

N(a,n,s): the number of live births of infants of sex *s* during year *n* and whose mother was born in year *n-a*;

M(a,n,s): the number of persons entering metropolitan France minus the number of people leaving metropolitan France during year *n*, of sex *s* and born in year *n-a*. This is the net migration in year *n*, for persons of sex *s* and born in year *n-a*.

To simplify the subsequent notations, we define *P(0,n,s)* as the number of live births in year *n* of babies of sex *s*. Furthermore, *D(0,n,s)* and *M(0,n,s)* are well defined by the above description and correspond respectively, for each year *n* and sex *s*, to the number of deaths of babies born in year *n* and to the number of newborns entering the country minus the number of newborns leaving the country. It will be assumed that the ages of women at childbirth are between 15 and 55 years inclusive, meaning that *N(a,n,s)*=0 for $a \leq 14$ and $a \geq 56$.

In addition, populations at risk are defined for deaths and births. Populations at risk are counted in person-years and depend on the number of people observed, but also on the period of time over which these people are present. For deaths, this corresponds to the population on 1 January for the year in question, plus half of net migration (assuming that inflows and outflows are evenly distributed throughout the year).

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

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

where $a = 0$.

For births, the number of person-years at risk is the average number of women during the year in question, assuming that migration flows and deaths remain uniform:

$$R_N(a,n) = P(a,n,women) + 0.5M(a,n,women) - 0.5D(a,n,women).$$

We also note $M(n) = \prod_{a,s} M(a,n,s)$, $N(a,n) = N(a,n,girls) + N(a,n,boys)$ and $N(a,n) = \prod_{a} N(a,n,s)$.

When noting normal distributions, we will indicate the standard deviation (rather than the variance).

3.1. Migration

The total net migration is directly projected using a first-order autoregressive model, where M_{lt} represents the long-term net migration and ε_M represents white noise:

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

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

In order to ensure a stationary process, the constraint $|\rho_M| \leq 1$ is imposed. This modelling reflects the fact that it is estimated that net migration will continue to be stable on average and will oscillate around a long-term trend. The amplitude of possible future oscillations is determined by past amplitudes. Furthermore, a very informative *a priori* is set with regard to the long-term trend by assuming, as was the case in the work of Blanpain & Buisson (2016a), that this can be estimated from the average net migration over the recent period, i.e. 80,000 persons. The *a priori* distribution for the long-term trend is therefore $M_{lt} \sim N$(80,000;

10,000). The parameters M_{lt}, ε_M, ρ_M and σ_M are estimated by means of Bayesian inference based on the net migration for the period 1995-2013.

To project the total net migration, the model parameters are randomly drawn 1,000 times according to their *a posteriori* distribution and for each set of parameters, the development of net migration is simulated according to the first-order autoregressive process. Once the net migration has been projected, it is broken down by sex and age in accordance with fixed rates calculated on the basis of the distribution of net migration by sex and age over the recent period and smoothed, as described in Blanpain & Buisson (2016a).

3.2. Mortality

As has already been mentioned, the logarithm of age-specific mortality rates appears to develop in a linear manner over time. Nevertheless, mortality rates develop at a different rate for each age over time. The number of deaths observed is directly modelled in accordance with Poisson's law, which is based on the mortality rate and the population at risk. The latter corresponds to the number of person-years present in metropolitan France in the year in question. Poisson's law is currently used to model a number of events occurring over a given period of time. It is often used to model the number of deaths in demographic work. The following model (developed by Bryant & Zhang, 2014) is used, where $\mu_D(a,n,s)$ corresponds to the mortality rate for year n for persons of sex s and age a:

$$D(a,n,s) \sim Poisson\left(\mu_D(a,n,s)\ R_D(a,n,s)\right)$$

$$\log\left(\mu_D(a,n,s)\right) =$$
$$\beta^0 + \beta_a^{age} + \beta_{a,s}^{age:sex} + \beta_{a,n}^{age:year} + \varepsilon_{D,1}(a,n,s)$$

$\varepsilon_{D,1}$ are independent and identically distributed error terms according to centred normal distribution and standard deviation $\sigma_{D,1}$. The parameter β_0 is a constant, the parameter β^{age} gives the average age distribution of the logarithm of mortality rates. Finally, there are two terms that cross two dimensions: $\beta^{age:sex}$, which allows the specific effect of sex to be estimated for each age and $\beta^{age:year}$, which is a time effect specific to each age. It should therefore be noted that the development of the logarithm of age-specific mortality rates over time is the same for both women and men, since no term that crosses the dimensions of year and sex has been specified. This is because we wanted to limit the number of parameters to be estimated. When a term crossing the year-sex dimension was introduced, it was found that the *a posteriori* distribution was not correctly estimated due to a non-convergence of the Markov chains. At a third level, some of the parameters are modelled by means of dynamic linear models. For the $\beta^{age:year}$ parameter, this allows the development over time to be broken down, by age, into a level ($\theta^{age:year}$) and a trend ($\delta^{age:year}$):

$$\beta_{a,n}^{age:year} = \theta_{a,n}^{age:year} + \eta(a,n)$$
$$\theta_{a,n}^{age:year} = \theta_{a,n-1}^{age:year} + \delta_{a,n}^{age:year} + \upsilon(a,n)$$
$$\delta_{a,n}^{age:year} = \delta_{a,n-1}^{age:year} + \omega(a,n)$$

The terms η, v and ω are independent error terms that follow centred normal distribution.

To project age-specific mortality rates into the future, once the *a posteriori* distribution of all the model parameters has been estimated, it is sufficient to generate new trend terms, followed by new level terms and finally new $\beta^{age:year}$ parameters, up to the desired horizon.

3.3. Fertility

For fertility, we chose to proceed in three stages. First of all, the TFR is projected according to a first-order autoregressive model. The UN uses the same method for its third stage of fertility change, on the assumption that the TFR tends towards 2.1 in all countries (Alkema *et al.*, 2010). When compared with the method used by the UN, we have chosen to also estimate the parameters of the model by means of Bayesian inference rather than by maximum likelihood. We therefore remain within an entirely Bayesian framework for all our estimates and projections. The model is as follows:

$$ICF(n) = ICF_{lt} + \rho_F\left(ICF(n-1) - ICF_{lt}\right) + \varepsilon_F(n)$$

where $ICF(n) = \prod_{a=15}^{55} \frac{N(a,n,girls) + N(a,n,boys)}{R_F(a,n)}$ is the total fertility rate in year n. As was the case for net migration, after estimating the Bayesian inference, we simulate 1,000 possible trajectories for the development of this index up to the desired horizon.

The second stage consists of projecting the age-specific fertility rates μ_F, independently of the projection of the TFR. As is the case with mortality, these are defined by modelling births by means of a Poisson process:

$$N(a,n) \sim Poisson\left(\mu_F(a,n)\ R_F(a,n)\right)$$

by way of a reminder, $N(a,n)$ corresponds to the number of births in year n, given by mothers born in year n-a. Following the method proposed by Bijak *et al.* (2015), which is based on the

Lee-Carter method, we then modelled the logarithm of the fertility rate as the sum of a fixed age effect, a time effect for which the intensity and direction are different for each age, and a generation effect:

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

The time effect κ and the generation effect γ change in accordance with the first-order autoregressive processes:

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

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

where the error terms ξ and ζ follow normal laws of zero expectation. Once again, all parameters are estimated by Bayesian inference in order to subsequently produce 1,000 fertility rate simulations for each age and each future year. These projected rates extend linear trends, although the parameters φ_1 and Ψ_1 may, if they are strictly smaller than 1, cause the time effect or the generation effect to cancel out in the long term. The estimates give an *a posteriori* distribution of φ_1 and Ψ_1, which are very close to 1. This results in the fertility rates becoming abnormally high for certain ages, which leads to TFRs that are much higher than those projected in the first stage.

The third stage involves then correcting the age-specific fertility rate for each year and aligning it to the TFR initially projected. In order to do so, we simply multiply all of the rates in a given year by a constant. Note that no constraint was added for the average age at childbirth, whereas Insee's projections retain a ceiling at 32 years old based on experts' opinion (see Blanpain, this issue).

Lifetime fertility is based on the fertility rates of a given generation of women. Like the TFR, it is a synthesis of fertility rates at different ages. However, unlike the TFR, which is a cross-sectional indicator, this is a longitudinal indicator and therefore requires the fertile life of an entire generation to be observed before it can be calculated. This therefore limits the number of observation points in the past. This is why we, like many other authors, decided to model and project the total fertility rate. Life expectancy is also a cross-sectional indicator.

3.4. Projections Using the Components Method

The components method makes it possible to develop the population from one year to the next by noting that the population on 1 January of a given year is equal to the population on 1 January of the previous year, plus the number of births that took place during the previous year, minus the number of deaths and plus net migration. This translates into the following equations:

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

if $a\geq 1$ and $P\left(0,n,s\right)=N\left(n,s\right)$.

The number of deaths and births are obtained each year by means of random sampling in accordance with Poisson's law (see models). In order to do this, the persons at risk must be identified for deaths and the women at risk in the case of births. We begin by calculating deaths for each age, with the exception of deaths among newborns. We then deduce the women at risk for each age between 15 and 55 years (in order to do so, we need to know the figures for net migration and the number of deaths). Finally, we calculate the number of deaths among newborns. The distribution of the number of births in a given year between male and female is determined by the sex ratio, which is set at 1.05 in accordance with past observations.

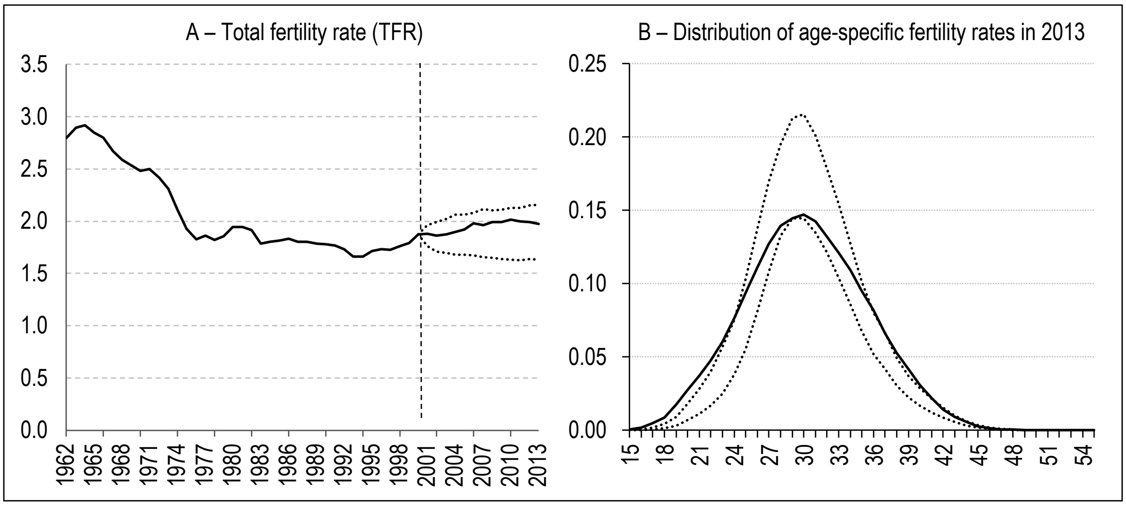
3.5. Validation of the Models

One way of testing the models used is to separate the data relating to the past into two categories: one part, approximately two-thirds, is used to estimate the models and the remaining part, approximately one-third, is used to compare the model estimates with reality.

In the case of mortality, we decided to estimate the model for the period from 1962 to 1995 and to compare the results during the period from 1996 to 2013. For fertility, we estimated the models over the period from 1975 to 2000 and we compared the results from the period between 2001 and 2013. It is clear that the logarithm of mortality rates is projected adequately at older ages (from around 35-40 years of age), but that the model used presents decreases in these rates that are much slower than what is actually observed. This is because, at very young ages, the logarithm of mortality rates is not linear, but is instead slightly concave. Moreover, mortality rates for young adults more or less stagnated in the 1980s and 1990s, before falling sharply. The model was not able to predict this sudden drop.

As regards fertility, the TFR observed is well within the 95% confidence interval of the probabilistic projections of the TFR. However, when we look at the distribution of age-specific fertility rates, it becomes apparent that the method used leads to a tighter distribution than is actually observed (Figure VIII). The deformation of the distribution of age-specific fertility

Figure VIII – **TFR and age-specific fertility rates, observed (1962-2013) and projected (2001-2013)**



Notes: The dotted lines indicate the 2.5% and 97.5% quantiles of the probabilistic projections and the solid line indicates the actual TFR and fertility rates (from 1962 to 2013).
Sources and coverage: Insee, population estimates and civil registry statistics (fertility rates), Metropolitan France. Author's calculations (probabilistic projections)

rates is therefore a little too pronounced in our projections.

4. Results of Bayesian Probabilistic Projections for France up to 2070

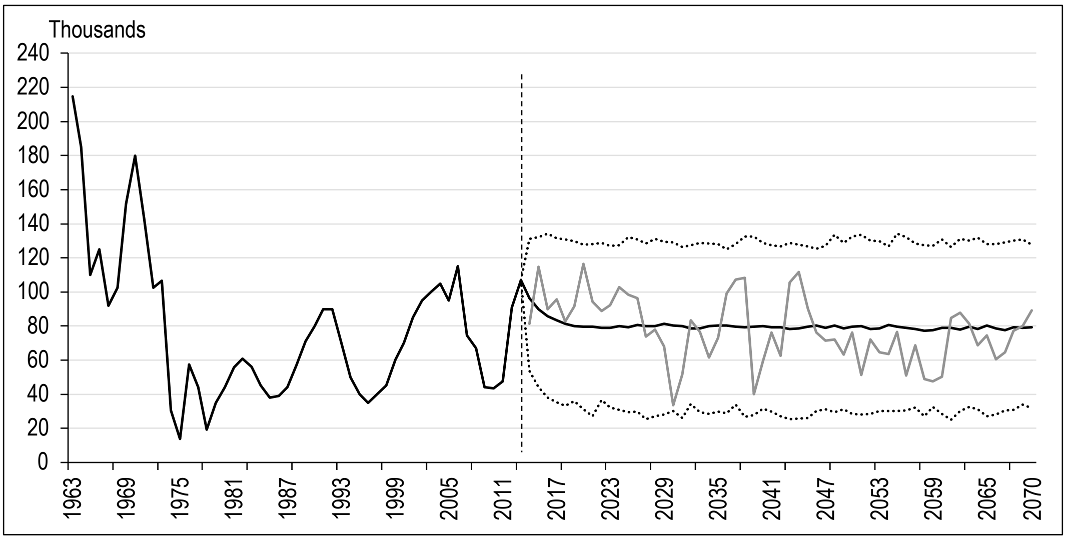
The parameters of the models for net migration, mortality and fertility were estimated by Bayesian inference using the open source software, *Stan* and the R *demest* package published by the Statistical Institute of New Zealand.[3] We simulated 1,000 values for each of these parameters according to their *a posteriori* law. We then generated 1,000 possible evolution trajectories for net migration, sex-specific and age-specific mortality rates and age-specific fertility rates. In the end, 1,000 estimates can be obtained for any demographic indicator derived from these three components, including the size of the total population. Confidence intervals of 95% or 80% are then derived from these, which contain 95% or 80% of the estimates, respectively.

4.1. Migration Projections: A Strong and Constant Uncertainty

Projected net migration follows a stable trajectory as this was specified in the model. The median of the 1,000 possible trajectories decreases in

3. *https://github.com/StatisticsNZ/demest*

Figure IX – **Net migration, past and projected**



Note: The dotted lines indicate the 2.5% and 97.5% quantiles and the solid line indicates the median of the *a posteriori* distributions. The light grey curve represents one of the 1,000 simulations.
Sources and coverage: Insee, population estimates and civil registry statistics (1962-2013), author's calculations (2013-2070); Metropolitan France.

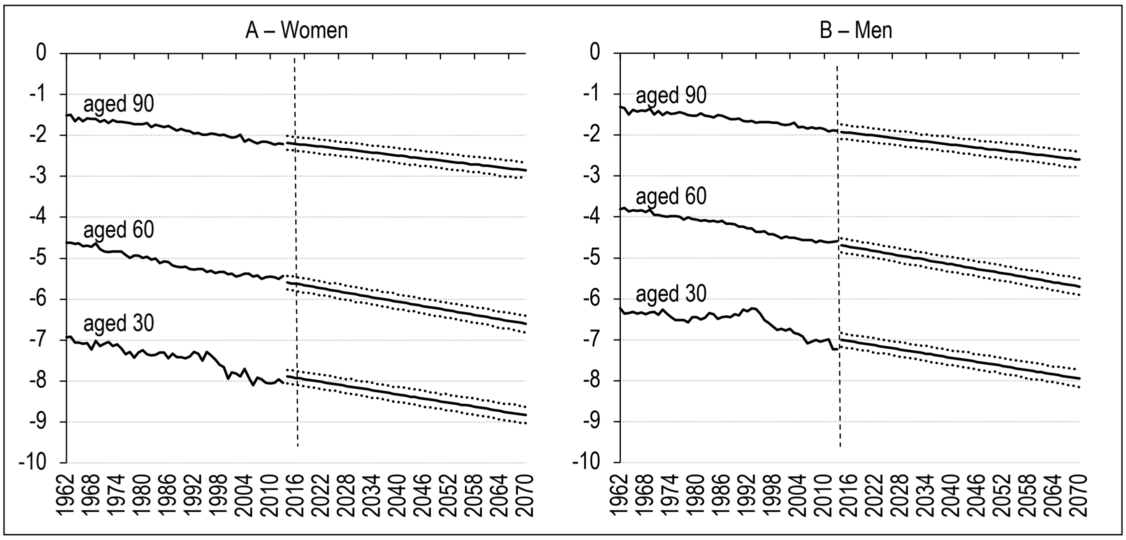
the first few years of projections before rapidly stabilising at 79,000 (Figure IX). The confidence interval also remains constant over time: at a probability of 95%, net migration will remain at between 29,000 and 129,000 each year. This amplitude is due to the significant fluctuations observed in the past and slightly exceeds the minimum and maximum observed in 1996 and 2006 respectively.

4.2. Mortality Projections: Little Uncertainty Given Past Developments

The model for mortality predicts that age-specific mortality rates will continue to decline in a linear manner, following the same trend for both males and females (Figure X). The uncertainty in the projected mortality rates does not increase over time. This is because the variance in the level and trend errors v and ω is very small compared with the variance in the error term η. Errors therefore do not accumulate over time. This is due to the fact that the trends observed are highly linear.

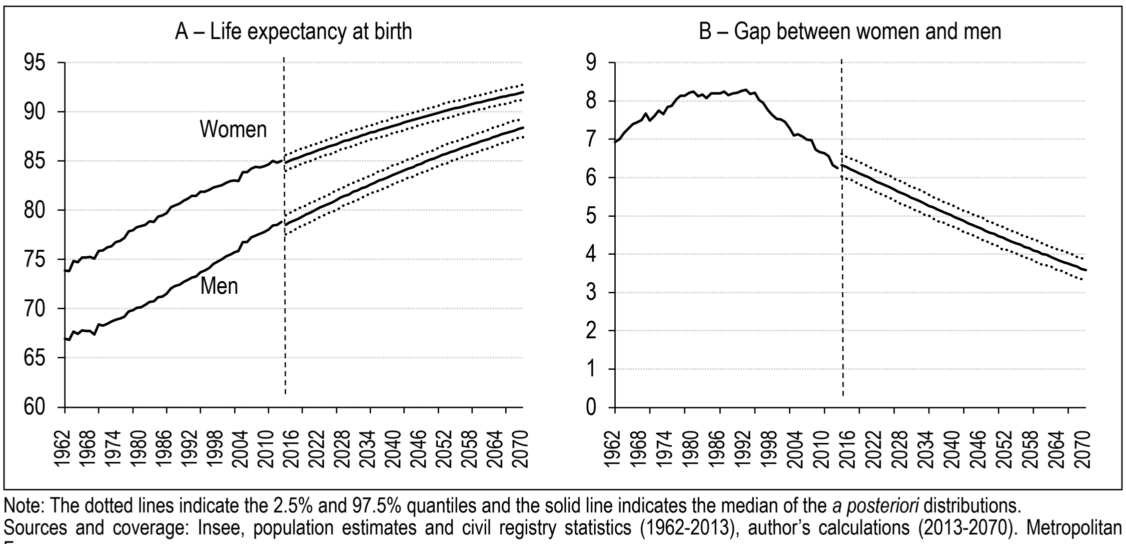
Due to the constant reduction in mortality rates, life expectancy will continue to increase in the coming years for men and women alike. The results of the model indicate that, with a probability of 95%, life expectancy at birth in 2070 will be between 91.2 and 92.8 years for women and between 87.4 and 89.4 years for men (Figure XI). The gap in life expectancy between

Figure X – **Changes in the logarithm of age-specific mortality rates, estimated and projected**



Note: The dotted lines indicate the 2.5% and 97.5% quantiles and the solid line indicates the median of the *a posteriori* distributions.
Sources and coverage: Insee, Metropolitan France. Author's calculations.

Figure XI – **Estimated and projected changes in women's and men's life expectancy and gender gap in life expectancy**



Note: The dotted lines indicate the 2.5% and 97.5% quantiles and the solid line indicates the median of the *a posteriori* distributions.
Sources and coverage: Insee, population estimates and civil registry statistics (1962-2013), author's calculations (2013-2070). Metropolitan France.

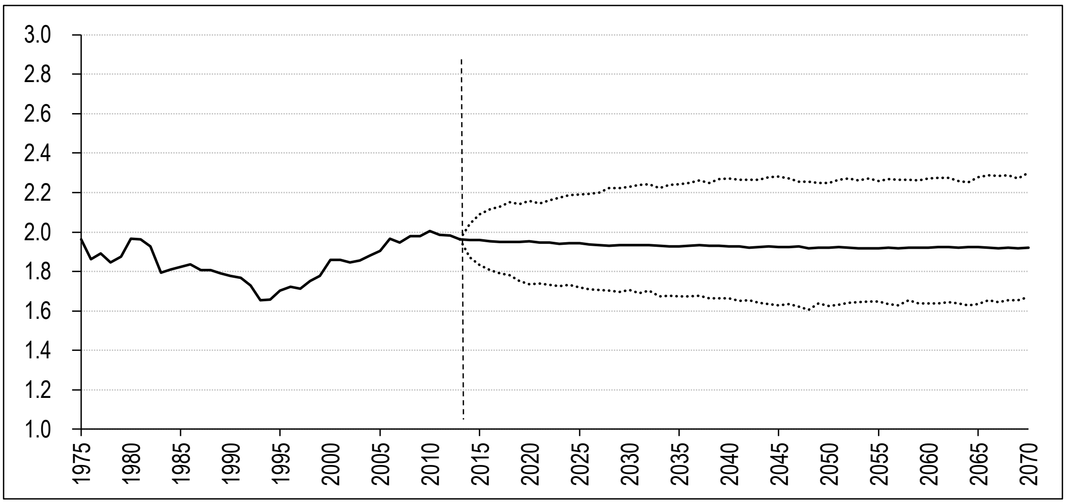
women and men will likely continue to narrow to reach 3.6 years in 2070 (between 3.3 and 3.9 years with a probability of 95%).

4.3. Fertility Projections: Births to Older Mothers and More Symmetrically Distributed Around the Modal Age

The median long-term TFR is 1.93, slightly below the mean of the *a priori* distribution, which is set at 1.95 (Figure XII). According to the model used, the TFR will be between 1.63 and 2.26 children per woman in 2070 at a probability of 95%. Unlike the projections for net migration and mortality rates, the confidence interval at 95% becomes wider over time. The uncertainty with regard to future fertility therefore becomes higher, in spite of having set a long-term TFR in the model.

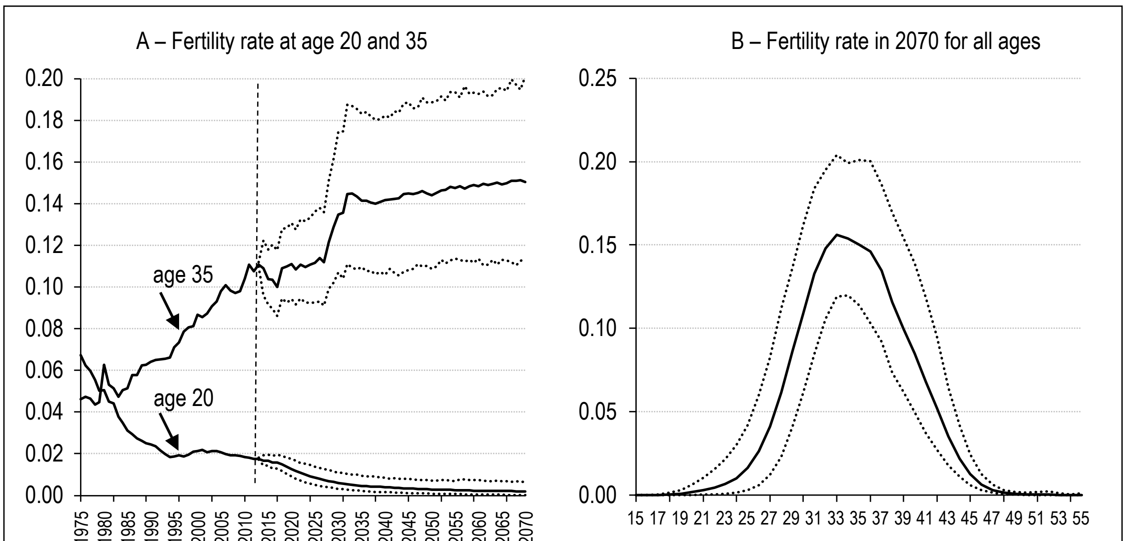
The age-specific fertility rates begin to stabilise from 2050 onwards (Figure XIII). The average age at childbirth rises rather quickly until around 2040, after which the increase continues but at a slower and slower rate until it reaches a value of between 32.2 and 35.9 years in 2070 (confidence interval of 95%). The distribution of age-specific fertility rates therefore shifts more and more to the right and becomes increasingly symmetrical, as evidenced by the changes in the measure of skewness, the median of which is tending towards 0 (Figure XIV).

Figure XII – **Changes in the total fertility rate, estimated and projected**



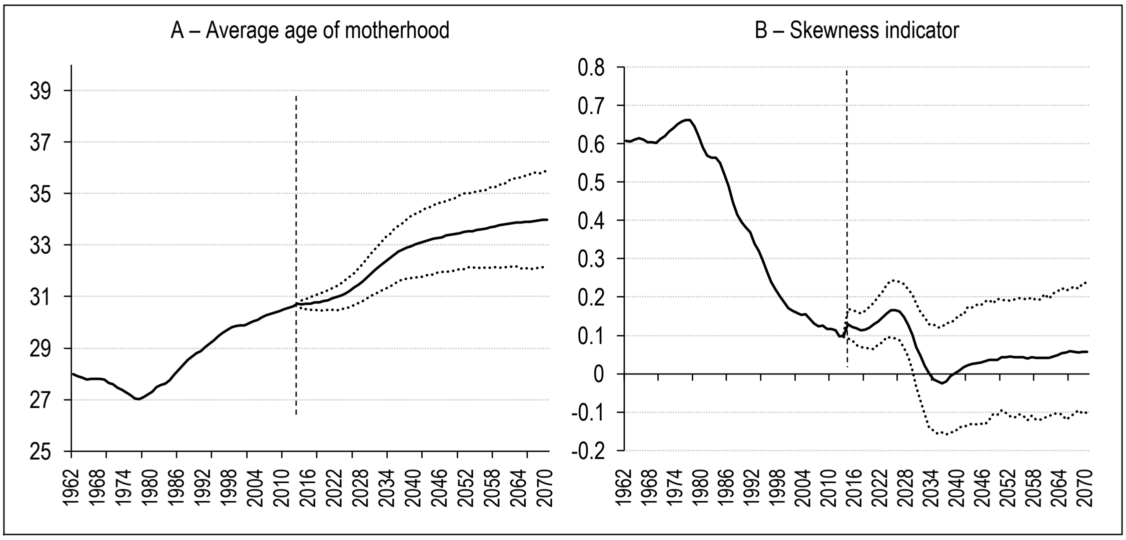
Note: The dotted lines indicate the 2.5% and 97.5% quantiles and the solid line indicates the median of the *a posteriori* distributions.
Sources and coverage: Insee, population estimates and civil registry statistics (1962-2013), author's calculations (2013-2070); Metropolitan France.

Figure XIII – **Changes in the fertility rates, estimated and projected**



Note: The dotted lines indicate the 2.5% and 97.5% quantiles and the solid line indicates the median of the *a posteriori* distributions.
Sources and coverage: Insee, Metropolitan France. Author's calculations.

Figure XIV – **Changes in the average age of motherhood and skewness of the age-specific distribution of fertility rates**



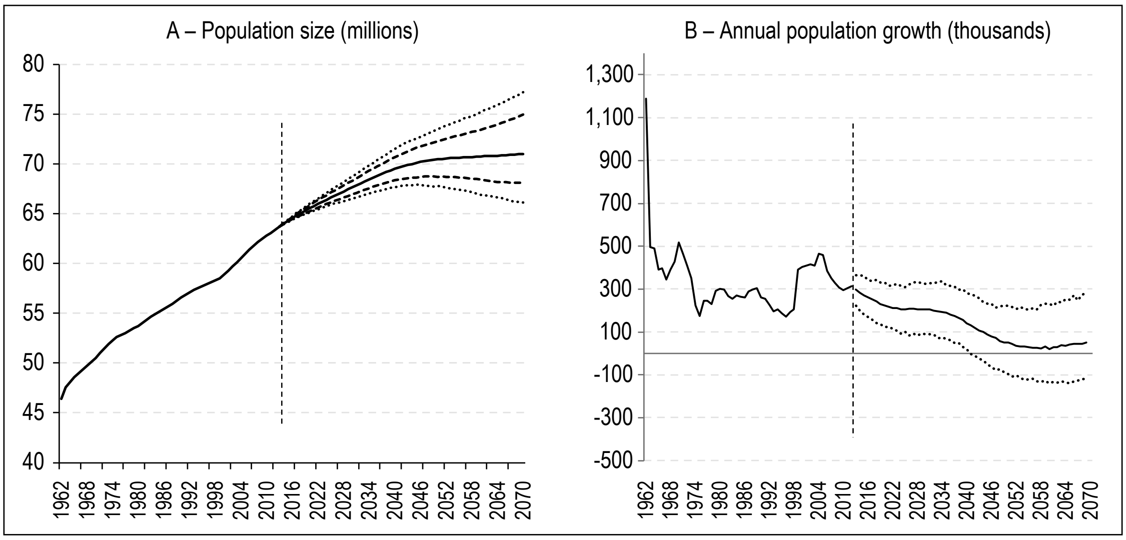
Note: The dotted lines indicate the 2.5% and 97.5% quantiles and the solid line indicates the median of the *a posteriori* distributions.
Sources and coverage: Insee, population estimates and civil registry statistics (1962-2013), author's calculations (2013-2070); Metropolitan France.

4.4. Total Population Projections: Growth Likely to Be Strong Until 2040 and Much Weaker Thereafter

The total population of metropolitan France will continue to grow until it reaches a level of between 66.1 million and 77.2 million in 2070 with a probability of 95%, and between 68.1 million and 75.0 million with a probability of 80% (see Figure XV). The median projection corresponds to a level of 71.0 million inhabitants in 2070. The population of metropolitan France could therefore increase continuously throughout the next fifty years, or it could increase before beginning to decline around 2050. According to the model used here, there is a 1% probability that the population will start to decrease from 2040 onwards (i.e. the population will reach its peak in 2040) and a 19% probability that this will occur in 2050. The uncertainty regarding the size of the population according to the model used is relatively minor until around 2040-2050, after which it increases more rapidly in the years that follow.

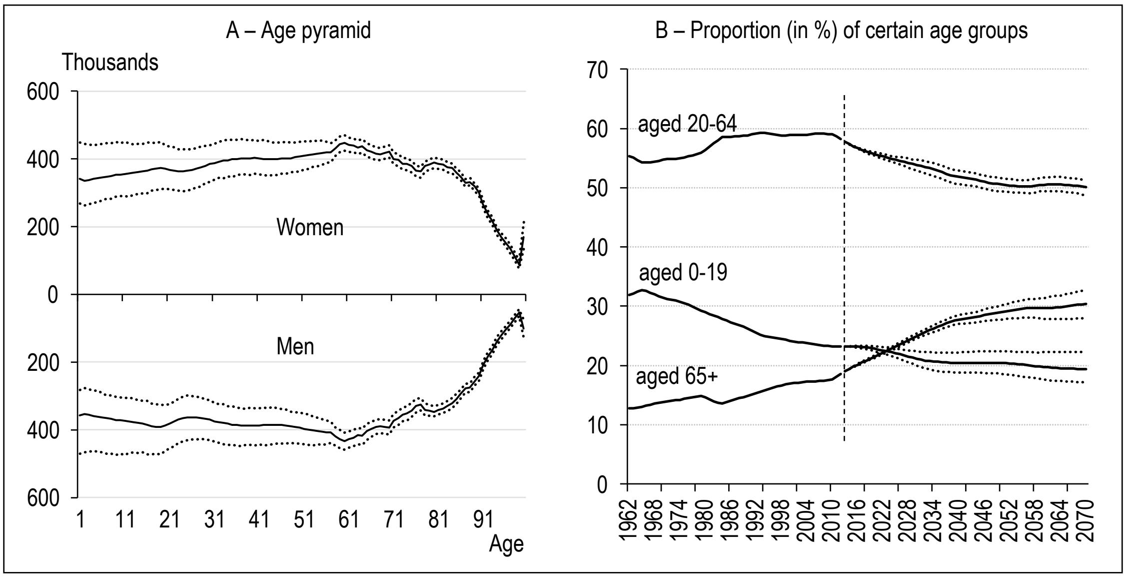
The structure of the population will also change, as can be seen in the population pyramid for 2070, the base of which is much straighter and thinner than the pyramid depicting current ages. The proportions of certain age groups will therefore decrease, particularly the youngest

Figure XV – **Past and projected changes in the total size of the population and annual population growth**



Note: The dotted lines indicate the 2.5% and 97.5% quantiles, the dashed lines indicate the 10% and 90% quantiles and the solid line indicates the median of the *a posteriori* distributions.
Sources and coverage: Insee population estimates and civil registry statistics (1962-2013), author's calculations (2013-2070); Metropolitan France.

Figure XVI – **Age pyramid for 2070 and changes in the proportion of certain age groups**

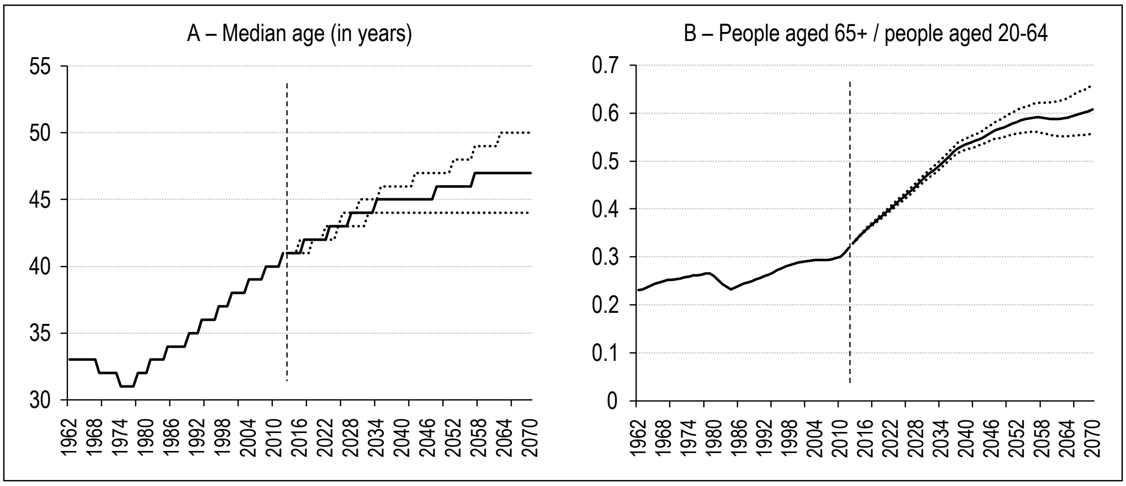


Note: The dotted lines indicate the 2.5% and 97.5% quantiles and the solid line indicates the median of the *a posteriori* distributions.
Sources and coverage: Insee population estimates and civil registry statistics (1962-2013), author's calculations (2013-2070); Metropolitan France.

(see Figure XVI): the proportion of people aged 0-19 years will continue to decrease slowly until it reaches a median level of 19% in 2070; 20-64-year-olds will follow the same pattern, with a median level of 50% in 2070. Conversely, the proportion of the population aged 65 and over will probably continue to increase until it exceeds the share of people aged under 20 in 2070. This figure increased from 13% in 1962 to 19% in 2013 and has the potential, with a probability of 95%, of making up between 28% and 33% of the population in 2070.

The population will therefore continue to age. The median age of the population, which was 41 years in 2013, could, with a probability of 95%, be between 44 and 50 years in 2070. As a result, the ratio of people aged 65 and over to people aged 20 to 64 years is likely to rise sharply in the coming years. The rapid and linear increase in this ratio between now and the early 2040s is largely due to the ageing of the large generations born during the baby boom. In fact, people born at the start of the baby boom in 1946 turned 65 in 2011 and those born at the end of the baby boom in 1975 will turn 65 in 2040. According to the models used, the ratio of those aged 65 and over to those aged 20-64 years, which today stands at 0.33, will reach a value of between 0.56 and 0.67 in 2070 with a probability of 95% (see Figure XVII).

Figure XVII – **Changes in the median age of the population and the ratio of people aged 65 and over to people aged 20-64 years**



Note: The dotted lines indicate the 2.5% and 97.5% quantiles and the solid line indicates the median of the *a posteriori* distributions.
Sources and coverage: Insee population estimates and civil registry statistics (1962-2013), author's calculations (2013-2070); Metropolitan France.

These probabilistic projections can be compared with the deterministic projections made by Insee. The deterministic projections concerning metropolitan France only cover the period from 2013 to 2050.[4] The central scenario adopted leads to a population that is slightly larger than the median of our probabilistic projections: according to the first projection, the population of metropolitan France would reach 71.7 million inhabitants in 2050 and 70.5 million according to the second. Furthermore, the confidence interval estimated by the probabilistic projections is much lower than the interval between the high and low population scenarios, which are the extreme scenarios of the deterministic projections. The difference between the two extreme deterministic scenarios is 11.1 million inhabitants in 2050, whereas the confidence interval of the deterministic projections for that same year is 5.7 million for the 95% confidence interval and 3.6 million for the 80% confidence interval.

4.5. Discussion

According to the models described in this article and the simulations carried out, the population of metropolitan France is expected to continue increasing in the coming decades. However, there is a non-negligible probability that it will start to decline before 2070, although this is less likely than an increase or stabilisation. The structure of this population is also likely to change: a general ageing of the population is expected due to increased life expectancy, a stagnating trend in the TFR and the continued arrival of baby boomers at retirement ages. The model used to project net migration is the simplest of the three models used. The lack of age-specific data on people entering and leaving the country precludes the use of Poisson modelling to obtain the rates, as we did for the number of deaths and the number of births. In general, models for projecting net migration are less sophisticated and have been the subject of less research effort than those for mortality and fertility, since the available data are less rich. Nevertheless, it is worth noting that some countries, such as New Zealand in particular, which have detailed data concerning people entering and leaving the country, are starting to offer advanced modelling of migration phenomena, taking account of a large number of parameters, such as the level of education attained by the population (Bryant & Zhang, 2014). Since our modelling is fairly simple, it follows that most of the past changes in net migration are considered noise. Since this noise is then propagated into the future, the confidence intervals of the projected net migration are very wide and therefore reflect our level of uncertainty about the future evolution of migration. This is why we have restricted the estimation of the parameters of the model (and therefore the variance of the error term in particular) to the period 1995-2013, to ensure that we do not take account of large fluctuations in the migratory balance that are too old. Estimating the model over a longer period would have led to even greater uncertainty about the future development of net migration.

Unlike net migration, mortality trends are very stable and the model used is able to take account of these trends without considering them to be predominantly noise. As a result, the confidence intervals of projected mortality rates and life expectancy are very small. This may seem misleading, as one could be led to believe that we are almost certain of what will happen. In reality, it is important to remember that the confidence intervals on future mortality levels are conditionally determined by the model taking the correct approach to reality. Indeed, such levels of confidence can only be attained for future mortality rates by assuming that the observed trends will continue. In spite of this, the model used does not take account of certain peculiarities of mortality in France. Firstly, it does not allow gender-specific changes in the logarithm of the mortality rate to be projected at a given age. Furthermore, it appears that generations born after the Second World War have very little gain in terms of mortality at a given age when compared with previous generations, regardless of the age in question (Blanpain & Buisson, 2016a). The model used does not allow such generation effects to be taken into account: deviations from the general trend are therefore treated as noise and included in the error terms rather than being seen as a well-identified effect. The resulting projected life expectancies are therefore somewhat lower than those obtained by the projections made by Blanpain & Buisson (2016a).

Fertility is modelled differently from net migration and mortality. In fact, unlike mortality rates, fertility rates do not evolve in a regular manner over time. They can increase and then decrease or vice-versa, and therefore intersect. Extending fertility rates in accordance with linear trends also leads to situations that appear implausible in the light of other fertility indicators, such as the TFR and the peak fertility rate attained during the year, which have remained more or less stable

4. *see https://www.insee.fr/fr/statistiques/2859843*

since 1975. The idea was to initially extend the TFR, which is an indicator that reflects the level of fertility, using the same method as was used to project net migration. We then extended the age-specific fertility rates in accordance with the method described by Wiśniowski *et al*. (2015), and we modified these rates to bring them back down to the TFR projected in the first place. This provides a fairly realistic trend of age-specific fertility rates, with the distribution shifting towards older ages and becoming more symmetrical. This approach (projecting an aggregated indicator and then breaking it down into detailed categories) is not new and is also the approach adopted by the UN. The disadvantage is that a long-term TRF must be set and the level chosen obviously affects the results.

* *
*

Probabilistic population projections provide new insights into possible population change. They make it possible, under certain modelling assumptions, to quantify the level of uncertainty concerning the future development of demographic indicators and in particular the evolution of total population size. They therefore offer a clear advantage over deterministic projections based on scenarios for which the probability of occurrence is not quantified. Any demographic indicator, whether it be life expectancy, the average age of motherhood or the proportion of people aged 65 and over, can be determined with some degree of probability. One of the potential difficulties in interpreting the results stems from the fact that one should not think in terms of a single point, but rather in terms of the probability distribution, just as a dice cannot be defined by just one of its six sides, even if it is loaded. Instead, it is by giving the probability that each number will appear that we will have a good description of the dice and what we can expect when it is rolled. Once this difficulty has been overcome, the interpretation and use of probabilistic population projections offers a great deal of freedom and flexibility. Conversely, the results of deterministic projections become complicated to use and disseminate when the number of scenarios under consideration is multiplied by the effect of several hypotheses intersecting.

There are a number of ways in which the methods used in this article, and therefore the results, can be improved. The first step is to better understand the migration phenomena by performing a detailed analysis of persons entering the country. It would also be interesting to look more thoroughly at estimates of flows of persons leaving the country, both now and in the past, which are relatively new in France, taking account of the available data. When projecting mortality, it would be useful to incorporate a generation effect and to allow mortality rates to develop differently for women and men. Several models are possible for this; however, the difficulty still remains that if there are too many parameters, there is a high risk that the model will not be identified or that the convergence of the Markov chains used to estimate the *a posteriori* distributions will be poor. In order to improve the projection of age-specific fertility, one could, as has already been done in several studies, find a parametric model of the distribution of age-specific rates. Although it would not necessarily be easy, it would then suffice to extend these parameters, as in the case of Lee-Carter modelling, by detecting regularities and trends in the development of these parameters. Beta distribution is a possible model, but its rounded shape would not represent the data well. Gamma distribution would better reflect the distribution of fertility rates, but it is defined on a support that is open to the right. It must therefore be truncated to ensure that there are no unrealistic results. The Hadwiger number presents a third option, as it seems better suited to modelling the distribution of fertility. The downside is that it can take a long time to estimate its parameters and their interpretation is not necessarily obvious. So why not propose an *ad hoc* function that faithfully reflects the observed data? It could be tempting to estimate the distribution of fertility rates in a non-parametric way, i.e. in reality by using a very large number of parameters. The difficulty then lies in the projection of these very many parameters. We could also consider developing structural models for the three components of population change that would make it possible to explain past change according to more detailed mechanisms and based on external variables; however, this would also require to have a sufficient number of elements to allow us to project the evolution of the variables. It would also be very informative to conduct sensitivity analyses, which would allow to test how the results vary when certain assumptions in the models are changed slightly. This would help to better understand and quantify the precise role of each component in population change.

As can be seen, there is undoubtedly much room for improvement, and this will require significant investments in research into understanding and modelling migration, mortality and fertility. This would only be beneficial for probabilistic population projections, the degree of uncertainty of which depends above all on our knowledge (or ignorance) of these topics. Finally, it is important that we do not compare probabilistic population projections with deterministic population projections. The latter remain extremely useful and allow us to test what would happen in the future in a given scenario. The general conclusions reached are also very consistent with those reached via deterministic projections of changes in population size and age structure. However, it is primarily up to the users of population projections to choose the approach that best suits them, depending on what they are using them for. Probabilistic and deterministic projections are two different ways of tackling uncertainty and trying to shed light on the future. ☐

BIBLIOGRAPHY

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*, this issue.

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 Zeland. *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 Officials 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

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 Zeland. https://www.stats.govt.nz/information-releases/national-population-projections-2016base2068

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 obstables à leur réalisation en France et en Italie. *Population-F*, 66(2), 401–432. https://doi.org/10.3917/popu.1102.0401

United Nations, Population Division (2017). World Population Prospects: The 2017 Revision, key findings and advance tables. *Working Paper* N° ESA/P/WP/248.

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

Evaluating Probabilistic Population Forecasts

Nico Keilman*

Abstract – Statisticians have developed scoring rules for evaluating probabilistic forecasts against observations. However, there are very few applications in the literature on population forecasting. A scoring rule measures the distance between the predictive distribution and its outcome. We review scoring rules that reward accuracy (the outcome is close to the expectation of the distribution) and sharpness (the distribution has low variance, which makes it difficult to hit the target). We evaluate probabilistic population forecasts for France, the Netherlands and Norway. Forecasts for total population size for the Netherlands and for Norway performed quite well. The error in the jump-off population caused a bad score for the French forecast. We evaluate the age and sex composition predicted for the year 2010. The predictions for the Netherlands received the best scores, except for the oldest old. The age pattern for the Norwegian score reflects the under-prediction of immigration after the enlargement of the European Union in 2005.

JEL classification: C15, C44, J11
Keywords: probabilistic population forecast, scoring rule, cohort component model

* *Department of Economics, University of Oslo, Norway (nico.keilman@econ.uio.no)*
Acknowledgments – Excellent comments from Laurent Toulemon and from three anonymous reviewers are gratefully acknowledged.

Received March 2019, accepted February 2020.

Citation: Keilman, N. (2020). Evaluating Probabilistic Population Forecasts. *Economie et Statistique / Economics and Statistics*, 520-521, 49–64. https://doi.org/10.24187/ecostat.2020.520d.2033

Most statistical agencies who compute population forecasts do so using a deterministic approach (NRC, 2000). They analyse historical trends in fertility, mortality, and migration, and extrapolate those trends into the future, using expert opinion and statistical techniques. These extrapolations reflect their best guesses. In addition to computing a likely development of population size and structure, many agencies also compute a high and a low variant of future population growth, in order to tell forecast users that future demographic developments are uncertain. For example, the previous official population forecast for France indicates 76.5 million inhabitants in 2070, if current trends continue (Blanpain & Buisson, 2016). However, population growth to 2070 might be weaker or stronger than what current trends suggest, leading to population sizes between 66.1 and 87.6 million persons. The forecasters assumed high and low trajectories for future fertility (1.8 or 2.1 children per woman on average after 2020), life expectancy of men (between 87.1 and 93.1 years in 2070) and women (between 90 and 96 years), and international migration (a migration surplus between 20,000 and 120,000 persons annually).

One important drawback of such a deterministic approach is that it fails to quantify uncertainty. We do not know if chances are 30, 60, or 90% that France in 2070 will have between 66.1 and 87.6 million inhabitants. Yet in many planning situations, it is important for the users to know how much confidence they should have in the predicted numbers. How robust should the pension system be with respect to fast or slow increases in life expectancies? Should we plan for extra capacity in primary schools, in case future births turn out to be much higher than expected? Indeed, as Keyfitz (1981) wrote almost 40 years ago: "Demographers can no more be held responsible for inaccuracy in forecasting population 20 years ahead, than geologists, meteorologists, or economists when they fail to announce earthquakes, cold winters, or depressions 20 years ahead. What we can be held responsible for is warning one another and our public what the error of our estimates is likely to be".

This is why the statistical agencies of some countries have started to publish their forecasts in the form of probability distributions, following common practice in, for example, meteorology and economics. Statistics Netherlands pioneered the field (see Alders & De Beer, 1998). Statistics New Zealand (2011) and Statistics Italy (ISTAT, 2018) are the other two known examples. In this connection, one should also mention the Population Division of the United Nations, which is responsible for regular updates of population forecasts for all countries of the world. In 2014, the Population Division issued the first official probabilistic population forecasts for all countries, using the methodology developed by Raftery *et al.* (2012).[1] The aim of a probabilistic forecast is not to present estimates of future trends that are more accurate than a deterministic forecast, but rather to give the user a more complete picture of prediction uncertainty.

Demographers in these statistical agencies could build on work and methods developed by demographers and statisticians since the 1980s. Two developments are noteworthy. First, early contributions applied an analytical approach, assuming a stochastic cohort component model, in which the statistical distributions for fertility, mortality, and migration parameters were transformed into statistical distributions for the size of the population and its age-sex structure. One needed strong assumptions, or derived only approximate expressions for the second moments of the age-sex distributions. Nowadays, a simulation approach is common. It avoids the simplifying assumptions and the approximations of the analytical approach. The idea is to compute several hundreds or thousands of forecast variants ("sample paths") based on random draws for the input parameter values of fertility, mortality, and migration. The simulation results are stored in a database. Keilman (2009) gives an example for France. A second methodological change is that from a predominantly frequentist approach to a Bayesian view of probability. In the frequentist tradition, the probability of an event is linked to its relative frequency of occurrence. In contrast, in the Bayesian approach a probability is interpreted as a person's subjective belief. It is particularly useful when models rely on expert opinions, and when one combines this kind of information with data. The change from a frequentist to a Bayesian approach in population forecasting was part of a more general trend towards "Bayesian demography", which started to gain popularity about ten years ago (Bijak & Bryant, 2016). The probabilistic UN forecasts mentioned earlier provide important examples of the Bayesian approach. Costemalle (this issue) applies this approach to the case of France.

1. *See also http://esa.un.org/unpd/wpp/Graphs/Probabilistic/POP/TOT/*

Once a probabilistic forecast has been published, its accuracy can be evaluated some 10-20 years later, when *ex post facto* observed data for population size and age structure have become available. However, accuracy assessment is difficult to carry out directly because it requires comparing a forecaster's predicted probabilities with the actual but unknown probabilities of the events under study. For that reason, statisticians have developed "scoring rules", also called "scoring functions". A scoring rule is a measure for the distance between the predicted distribution of a demographic variable, and the empirical value it actually turns out to have. Gneiting & Raftery (2007) and Gneiting & Katzfuss (2014) review the field. The score that one finds for a certain variable has no intrinsic meaning. Only in a comparative perspective, one can interpret the scores in a useful manner. This explains why scoring functions are frequently used in comparing two "competing" probabilistic forecasts.

Although the methodology around evaluation of probabilistic forecasts and scoring rules has been known for some time, there are very few applications of scoring rules to population forecasting. Shang *et al*. (2016) evaluated the accuracy of probabilistic cohort-component forecasts for the UK, and compared two forecasting methods. They used a scoring rule for prediction intervals. Shang (2015) and Shang & Hyndman (2017) evaluated interval forecasts for age-specific mortality rates of various countries, and used interval scores to select the best among several methods of mortality forecasting. Alexopoulos *et al*. (2018) employed interval scores to prediction intervals of age-specific mortality of England and Wales and New Zealand, and evaluated the predictive performance of five different mortality prediction models. All four papers use holdout samples to evaluate the probabilistic demographic forecasts. Genuine out-of-sample evaluation of probabilistic demographic forecasts has not been attempted before, to the best of our knowledge.

The aim of this paper is to show how methods for evaluating probabilistic forecasts developed elsewhere can be applied to probabilistic population forecasts. We present and apply scoring rules for prediction intervals, and for simulated samples of future population size and age structures. We illustrate the scoring rules using data for France, the Netherlands, and Norway, and compare probabilistic forecasts computed by different researchers. The comparisons serve three purposes. First, we investigate how fast the accuracy of a given probabilistic forecast changes with lead-time, i.e. when it looks further into the future. Second, we compare the accuracy of two ("competing") probabilistic forecasts for the same country. Finally, the relative performance of forecasts across countries is analysed.

Section 1 discusses the way the results of a probabilistic forecast are made available: as prediction intervals, or by means of a database. Section 2 presents a number of scoring rules and their characteristics. Empirical illustrations are given in Section 3. We evaluate various probabilistic predictions for total population size and the population pyramid of the three countries, then we conclude.

1. Publishing a Probabilistic Population Forecast

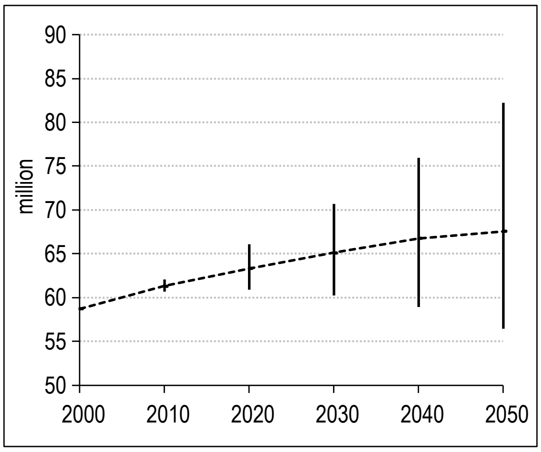
The methods one uses to evaluate a probabilistic forecast depend strongly on the way the forecast results are made available. There are two main methods. One is to publish prediction intervals for population variables. Alternatively, one may give the users access to a database with sample paths.

Costemalle (this issue) presents prediction intervals for the population of France, computed by a Bayesian approach. For instance, his Figure XV shows that there is an 80% probability that total population size in 2070 will be between 68.1 million and 75.0 million persons. The graph also shows 95% prediction intervals. These are much wider, because they cover more extreme situations. Other scholars (see Section 3 for examples) present their probabilistic forecasts in terms of 67% prediction intervals.

Figure I plots 80% prediction intervals for the population of France taken from the so-called UPE-project, to be discussed below. The jump-off year of this probabilistic forecast was 2003. Thus in 2050, 47 years into the future, the 80% prediction interval is 82.2 – 56.5 = 25.7 million persons wide. This is much wider than Costemalle 80% interval of 75.0 – 68.1 = 6.9 million persons (after 46 years). Different perceptions of prediction uncertainty for future fertility, mortality, and international migration lead to sharper (optimistic) or wider (pessimistic) prediction intervals.

These examples illustrate a more common finding, namely that different authors use different coverage probabilities for their prediction intervals. Selecting a coverage probability of 67 or 80% covers the majority of forecasts but excludes the more volatile tail of the error distribution. Those who use a coverage

Figure I – **Median values and 80% prediction intervals for total population of Metropolitan France**



Note: Median values on dashed line, 80% prediction intervals on solid lines.
Reading note: Chances are 50% that population size in 2050 will be less than 67.7 million; a population larger than 67.7 million is equally likely. There is an 80% probability that total population in 2050 will be between 56.5 and 82.2 million.
Sources: Keilman (2009).

probability of 95% do so, probably, because they have in mind a tradition in social science that implies constructing 95% confidence intervals or performing hypothesis tests with a low probability (5%) for type I errors (i.e. when a null hypothesis is rejected whereas in fact it is true). On the other hand, a prediction interval with coverage probability of 67 or 80% gives the user of the forecast an idea of how things might deviate from the point forecast. This is very different from constructing confidence intervals and from hypothesis testing. We will use both 67 and 80% prediction intervals in the empirical examples of Section 3.

Prediction intervals present only a summary of the complete probability distribution for the variable in question. Sometimes one can assume that the underlying distribution is approximately normal. In such cases, one can infer the parameters of the distribution from the upper and lower bounds of the interval. However, some population variables are restricted to a certain part of the real line, such as the share of the elderly in the population (between zero and one), and a normal distribution is not appropriate. In such cases one loses much information by publishing prediction intervals only, and not the underlying distributions.

The most of information becomes available when all simulated trajectories are stored in a database, to which the user has access (Alho & Spencer, 2005). A prominent example is the set of probabilistic population forecasts for 18 European countries, commonly known as the UPE-forecasts (“Uncertain Population of Europe”). The cohort-component model was applied 3,000 times for each country, with a deterministic jump-off population (as of 1 January 2003) and probabilistically varying values for age-specific fertility, mortality, and net migration. The forecast horizon was 2050. The UPE-forecasts have two attractive features. First, an explicit aim was to quantify uncertainty in such a way that it would reflect historical volatility in fertility, mortality, and international migration. Second, the project provided the first comprehensive look at empirical correlatedness of forecast errors in fertility, mortality, and migration across countries. More information, including a number of published and unpublished papers, is available at the UPE website.[2] The website contains a databank with simulation results (N=3,000) for men and women in five-year age groups in each country at ten-year (2010(10)2050) time intervals. This means that the user can build his or her histogram(s) for the variable(s) of interest. In Section 3, we will use the forecasts of the population pyramids for 2010 for France, the Netherlands, and Norway to illustrate the scoring rules discussed in Section 2.

2. Evaluation

Write the variable for which one computes a forecast as X, with cumulative distribution function (CDF) defined as $F(x) = P(X \leq x)$. The probability density function (PDF) of X is $f(x) = \frac{dF(x)}{dx}$. We assume throughout the existence of the integrals and various moments of the probability distributions. More fundamental treatments based on probability-theoretic considerations can be found in, for instance, Gneiting & Katzfuss (2014) and Gneiting & Raftery (2007). Write y for the observed value of X. A scoring function $S(F(x),y)$ assigns a numerical value (a “score”) to the forecast $F(x)$, given the observation y. $S(F(x),y)$ takes values in the real line $\mathbb{R}$ (including, possibly, plus and minus infinity).

A natural starting point for defining a scoring function is the following: a forecast that predicts the actual outcome with high probability should receive a good score. This works well for categorical forecasts, when X is a discrete random variable. However, we are dealing with forecasts for the number of persons (by age, sex, and forecast year), and X is closer to a continuous than a discrete random variable (unless the forecast is for a very small population). Henceforth we

2. *http://www.stat.fi/tup/euupe/index_en.html*

shall assume that the forecast and the scoring function are based on a continuous random variable. Many of the scoring functions start from the following two principles. First, an observation close to the median or the expected value of the predictive distribution gives a good score – the closer the better. In that case, the scoring rule is sensitive to distance (Staël von Holstein, 1970; Murphy, 1970). Second, given an observation, a narrow ("sharp") predictive distribution gives a good score – the narrower the better. For example, an 80% prediction interval that covers a certain observation represents a better forecast than an equally wide 67% interval that covers the same observation, because it is relatively difficult to hit the target when the PDF has low variance. However, the two principles are not equally important. One may argue that when the observation is "too far" from the median or expected value, one should no longer reward a narrow PDF. In other words, if the forecaster "takes a chance" (i.e. predicts a narrow PDF), the forecast should have a good score when the forecast is close to the median or expected value, but not when it is too far away. What one means by "too far away" is unclear, and it differs between scoring rules. The example above puts it as "the observation falls outside the prediction interval". This choice may be criticized: it rests on an extremely sharp dichotomy. In a very small interval around the upper or the lower bound of the prediction interval, the forecast changes very abruptly from having a good score to being punished for having an observation just outside the interval. To put it differently, given the predictive distribution and the observation, a prediction interval with the lower bound slightly lower than the observation gives a good score, whereas a bad score arises when the lower bound is slightly higher than the observed value. Coverage probabilities are arbitrary (80% is often used, but 81% or 79% work equally well). Therefore, one should be careful when defining the notion of "too far away".

Some of the scoring rules that we will discuss below indeed follow the idea that closeness is more important than sharpness. However, as we shall see, what we mean by "too far" is different for different scoring rules. Other scoring rules treat the two principles as independent. We say that a scoring function is negatively oriented when a lower score implies a better forecast, and the other way round for a positively oriented scoring function. Hence, a negatively oriented scoring function may be interpreted as a cost function, whereas a positively oriented scoring function reflects rewards.

Many different scoring rules have been proposed, depending on the nature of the forecast. Gneiting & Raftery (2007) and Jordan *et al*. (2019) give extensive overviews of the field. We will restrict ourselves to scoring rules for continuous random variables. One class of scoring rules applies to density forecasts based on closed-form expressions of the CDF or the PDF. An example is the logarithmic score $\text{LogS}(F(x),y) = -\log(f(y))$. A different class of scoring rules, more appropriate for the subject of this paper, evaluates simulated samples – in that case, the predictive distribution is not available analytically. A second distinction is that between univariate forecasts and multivariate forecasts. In the latter case, both the predicted variable X and the observation y consist of a vector. Jordan *et al*. (2019) developed the scoringRules package for R, which covers a wide range of situations in applied work.

Below we will introduce three types of scoring rules: those based on the first two moments of the predictive distribution only (Section 2.1), those stemming from the simulated complete predictive distribution, available as a sample (Section 2.2), and finally those for which one only has prediction intervals (Section 2.3).

2.1. Variance-based Scoring Functions

Assume a unimodal PDF of the forecast. When the actual outcome is close to the centre of the predicted density (as characterized by the mean, the mode, or the median), this forecast is better than one for which the outcome is far away from the centre. Stated differently, when there is little variation in X around y, the forecast scores better than when there is much variation. This leads to a variance-based scoring function, written as VS henceforth, and defined as follows.

Let VS be the variance of X around the observed value y, or

$$VS = \int (x-y)^2 f(x)\,dx \qquad (1)$$

For y equal to the expectation of X (written as μ), VS reduces to the variance of X, written as σ^2. Expression (1) leads to

$$VS = \sigma^2 + (\mu - y)^2 \qquad (2)$$

This defines a simple variance-based scoring function, which one could use to assess the quality of a unimodal predictive PDF. Gneiting & Raftery (2007) list it as a scoring function that corresponds to the so-called predictive model choice criterion or PMCC. One may apply it for analytical density functions as well as simulated samples. In the latter case, one uses estimates of σ^2 and μ from the sample. This scoring function is negatively oriented: a lower score indicates a better forecast. It rewards both accuracy – when y coincides with μ, the forecast is of optimal quality – and sharpness – a small variance gives a good score, irrespective of how far off the forecast was.

For a deterministic (point) forecast, σ^2 is zero and the forecast is μ. In that case, *VS* reduces to the squared error of the forecast. Errors of this kind form the basis of the Mean Squared Error frequently used in evaluations of deterministic population forecasts (Alho & Spencer, 2005; Smith *et al*., 2001; Keilman, 1990).

An alternative scoring function, also based on the first two moments of the predictive distribution, is the Dawid-Sebastiani score (e.g. Gneiting & Katzfuss, 2014)

$$DSS = \ln(\sigma^2) + (\mu - y)^2/\sigma^2 \qquad (3)$$

This scoring function is similar to the variance-based score *VS* in expression (2), but it gives different weight to the forecast variance σ^2.

A low variance leads to a good (low) score as long as $\frac{dDSS}{d\sigma^2} = \frac{1}{\sigma^2} - \frac{(\mu - y)^2}{\sigma^4} > 0$, or $\sigma > |\mu - y|$. Whereas *VS* always rewards predictive distributions with low variance, *DSS* does so if the observation y is less than one standard deviation away from the expectation of the predictive distribution.

Imagine a forecaster, who knows that her probabilistic forecast in due time will be evaluated by the scoring rules (2) or (3). Assume that at a certain stage of the production process of the forecast, the issue is to calibrate the forecast model. Use of scoring rule (2) or (3) implies that this calibration should focus on selecting an appropriate value for the mean μ of the predictive distribution – not the median or any other parameter of location. Indeed, there is a close relationship between model calibration and forecast evaluation. The situation is clear when there is only one user. However, things become more complicated when there are many users with different scoring rules (or with unknown scoring rules).

2.2. The Continuous Ranked Probability Score (*CRPS*)

The continuous ranked probability score (*CRPS*) might serve as a standard score in evaluating probabilistic forecasts of real-valued variables (Gneiting & Raftery, 2007). It is defined in terms of the predictive CDF $F(x)$ as

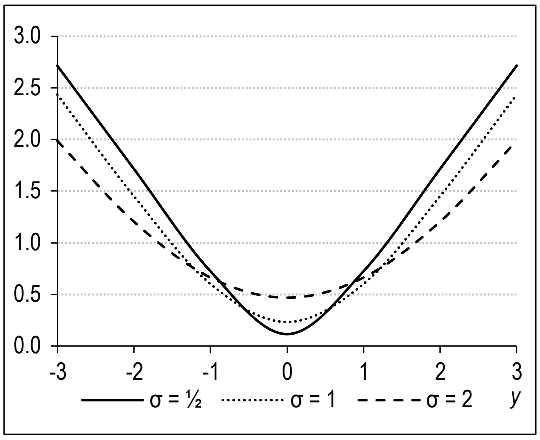
$$CRPS(F, y) = \int \left(F(z) - \mathbb{I}\{y \leq z\}\right)^2 dz \qquad (4)$$

where $\mathbb{I}\{y \leq z\}$ denotes the indicator function which is one if $y \leq z$ and zero otherwise. The particular form of the *CRPS* originates from the Brier score (Brier, 1950). The Brier score or probability score (PS) is a mean squared error of a categorical probability forecast. Murphy (1970) adapted it to the case of ordered categories for X, which led to the Ranked Probability Score or RPS. Matheson & Winkler (1976) proposed a RPS for the case of a continuous random variable, the *CRPS*.

Readily computable solutions to the integral above are few. Jordan *et al*. (2019) list the known cases. For instance, when $F(z)$ is the standard normal distribution $\Phi(.)$ with density $\varphi(.)$, $CRPS(\Phi,y)$ equals $y(2\Phi(y) - 1) + 2\varphi(y) - 1/\sqrt{\pi}$. The normal distribution with general expectation μ and standard deviation σ gives $\sigma CRPS(\Phi,(y - \mu)/\sigma)$.

It is worth to analyse a few concrete cases of the *CRPS*. We take the example of a normal distribution and assume, without loss of generality, that μ equals zero. Figure II plots this *CRPS* as a function of y, in other words, its sensitivity to distance. We show three cases, namely standard deviations of ½, 1, and 2. By construction ($\mu = 0$), the curves are symmetric around zero. As we might expect, the best score arises when y equals zero. The score becomes worse when y increases in absolute value, in other words, when y is far from μ. Sharpness of the predictive PDF (a low standard deviation) is only rewarded within a certain y-interval around zero. For instance, a perfect forecast (y equal to zero) scores better for $\sigma = ½$ (*CRPS* = 0.1168) than

Figure II – ***CRPS* for a normal distribution with expected value *μ* equal to zero and observations *y* ranging from -3 to +3**



Sources: Author's calculations.

for $\sigma = 2$ ($CRPS = 0.4674$). However, the PDF with $\sigma = 2$ scores better than the one with $\sigma = ½$ for observations y larger than approximately 0.9 in absolute value. For low σ-values, the interval where sharpness is rewarded is shorter than for high values.

Probabilistic population forecasts are commonly computed as simulated distributions, and one cannot compute the integral in (4). In that case, a useful starting point is the fact that (4) can be written as

$$CRPS(F, y) = E_F |X_1 - y| - ½ E_{F,F} |X_1 - X_2| \quad (5)$$

where X_1 and X_2 are independent random variables with distribution F (Gneiting & Raftery, 2007). The *CRPS* measures how close the observation y one can expect to be to the predicted variable X, corrected for the expected distance between all possible pairs of values of X. The latter expected distance is small when the standard deviation of F is small. Other things being the same, an increase in the standard deviation leads to a better score. However, when the standard deviation changes also the first expectation $E_F|X_1 - y|$ changes. Whether this score rule always rewards sharpness, or only on a certain interval, remains an empirical issue.

The *CRPS* reduces to the absolute error when F is a deterministic forecast. Assume that we have a forecast available in terms of a simulated distribution. Then the CDF is $\widehat{F}_m(x) = \frac{1}{m}\sum_{i=1}^{m}\mathbb{I}\{X_i \le x\}$ where m is the size of the sample, and (5) becomes

$$CRPS(\widehat{F}_m, y) = \frac{1}{m}\sum_{i=1}^{m}|X_i - y| - \frac{1}{2m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}|X_i - X_j|.$$

Implementation of this expression is inefficient, because it is of computational order $o(m^2)$. A more efficient and algebraically equivalent representation is (Jordan *et al.*, 2019, p. 6)

$$CRPS(\widehat{F}_m, y) = \frac{2}{m^2}\sum_{i=1}^{m}(X_{(i)} - y)\left(m\mathbb{I}\{y < X_{(i)}\} - i + ½\right) \quad (6)$$

where $X_{(1)}, X_{(2)}, X_{(3)}, \ldots, X_{(m)}$, is the sorted simulated sample. The *CRPS* as defined in (6) is always positive, because each term in the sum is positive.

2.3. Interval Scores

Many probabilistic population forecasts are presented as interval forecasts, not as (simulated) probability distributions (see Section 1). Consider a central $(1 - \alpha)$ prediction interval, with lower and upper endpoints that are the predictive quantiles at levels $\alpha/2$ and $(1 - \alpha/2)$, respectively.[3] Write l and u for the lower and upper quantiles. Gneiting & Raftery (2007) define the following score function

$$(u - l) + \frac{2}{\alpha}\left[(l - y)\mathbb{I}\{y < l\} + (y - u)\mathbb{I}\{y > u\}\right] \quad (7)$$

Given α, the Gneiting-Raftery interval score (*GRIS* henceforth) rewards forecasts for narrow prediction intervals that capture the observation y: when two competing forecasts have different prediction intervals for a given α, the forecast with shortest prediction interval gets the best (lowest) score. However, a value of y outside the prediction interval gives a bad (higher) score. The penalty for missing the prediction interval is larger for small than for large α. *GRIS* can be readily applied to the prediction interval of a variable for different lead times: 1 year ahead, 2 years ahead, 3 years ahead, etc.

There are situations in which *GRIS* does not reward sharpness, even when the interval captures the realization. Assume two competing forecasts with the same prediction interval $[l,u]$ that have different coverage probabilities. For instance, one forecast attaches 67% probability to the prediction interval $[l,u]$, whereas the other one has a coverage probability of 80% for the same interval. The second forecast is sharper and should receive a better score when the observation y falls inside $[l,u]$. However, this is not the case, as *GRIS* is independent of α in this situation. To solve the issue, one may use a slightly modified version of *GRIS*, namely

$$GRISmod = \alpha(u - l) + \beta\left[(l - y)\mathbb{I}\{y < l\} + (y - u)\mathbb{I}\{y > u\}\right] \quad (8)$$

where $\beta > 0$ is a parameter that determines how fast the score deteriorates when the observation is further away from either the upper or the lower bound of the prediction interval. A high β-value incurs a larger penalty than a low value. *GRISmod* rewards sharpness both for fixed α and different prediction intervals, and for the situation where one has a fixed prediction interval but different values of α. When β equals two, *GRISmod* equals α*GRIS*. In case one uses a β-value equal to the probability α, *GRISmod* reduces to $\alpha(u - y)$ for $y < l$ and to $\alpha(y - l)$ when $y > u$.

As an alternative to using scoring functions for prediction intervals, one could check how often actual data fall within the intervals. For instance,

3. *Note that we assume that the two quantiles are known. In case we want to evaluate interval forecasts when the nominal coverage level is specified, but the quantiles on which intervals are based are not specified, one cannot employ the approach outlined here (Askanazi* et al.*, 2018).*

Raftery *et al*. (2012) validated their Bayesian method of forecasting populations for 159 countries by estimating the model based on data for the 40-year period 1950-1990 to generate a predictive distribution of the full age- and sex-structured population for the 20-year period 1990-2010. They compared the resulting 80% and 95% prediction interval distributions with the actual observations, and checked the proportion of the validation sample that fell within their intervals. These proportions were close to the nominal values of 80 and 95%; therefore, the authors concluded that their approach was satisfactory. One important drawback of this method is that one compares data and prediction intervals for many variables, for instance the population size for all 56 countries in Africa at a certain date. However, regional correlations for fertility, mortality, and/or migration imply that the 56 population sizes are not independent. One has less data than originally thought and observed proportions cannot be compared directly with nominal values (Alho & Spencer, 2005, p. 248).

2.4. Scoring Functions Used in the Empirical Applications

In Section 3, we use the *CRPS* in expression (6) to evaluate forecasts for which detailed simulation results are available. In case we only have prediction intervals, we use the variance-based score *VS* of expression (2), the Dawid-Sebastiani score (*DSS*) of expression (3), and the interval scores (*GRIS* and *GRISmod*) of expressions (7) and (8). For *GRISmod* we assume a value for the parameter β equal to the probability α that was used to define the interval. *VS* and *DSS* use the expectation and the standard deviation of the predictive distribution. Since only upper and lower interval bounds are available, we assume normality and take the expectation as the mean of the two bounds, while we estimate the standard deviation as half the width of the interval for 67% intervals, and as the interval width divided by 2.564 for 80% intervals.

Note that the scores depend on the scale of the variable X for which we have a predictive distribution (which is the same scale as that of the observation y). Hence, when we compare the scores of two population forecasts for countries with very different population sizes, the smaller population will receive the best score, irrespective of its accuracy. For a fair comparison, we need to account for population size. We have normalized *VS*, *DSS*, *CRPS*, *GRIS*, and *GRISmod* as follows:

- we divided *VS* by μ^2, i.e. the square of the expected value of the predictive distribution;
- we normalized *DSS* by subtracting $2\ln(\mu)$ from the original *DSS* value;[4]
- we divided *CRPS*, *GRIS*, and *GRISmod* by μ.

3. Findings

Below we illustrate the scoring rules mentioned in Section 2.4 by evaluating probabilistic population forecasts for three countries: France, the Netherlands, and Norway. We focus on total population size (Section 3.1) and on the population pyramid (Section 3.2) of the three countries. The data stem from various sources:

1. At the UPE-website (see Section 1), samples (N = 3,000) for the forecasts of the population pyramid for the three countries are available for the years 2010, 2020, …, 2050. We use results for 2010.

2. Alho & Nikander (2004) report 80% prediction intervals and medians for total population size, amongst others, for each year in the period 2004-2050 for all UPE-countries. We use results for 2004-2019.

3. For the Netherlands, we have information about the official probabilistic population forecast with jump-off year 2000 (see Statistics Netherlands, 2001). The tables give 67% prediction intervals and expected values for total population for each year during the period 2000-2050, and for five-year age groups of men and women at five-year intervals.

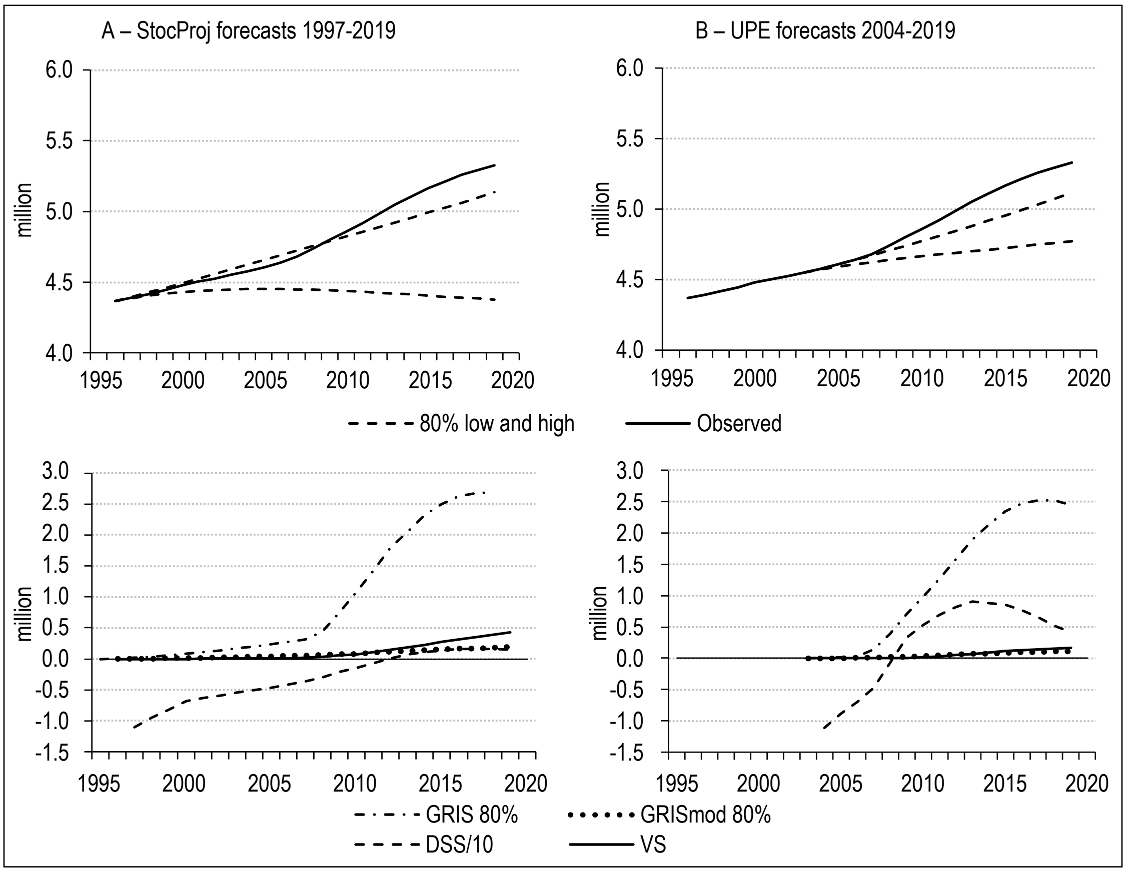
4. For Norway, we use results of the so-called StocProj ("Stochastic Projections") project (Keilman *et al*., 2002). The purpose was to compute a probabilistic population forecast with jump-off year 1996. Detailed simulation results are no longer available, but we use instead 80% prediction intervals for total population size for the years 1997-2019.

3.1. Population Size

Figure III shows our findings for Norway. There are four graphs, two for the StocProj forecast (left), and two for the UPE-forecast (right). The upper two graphs plot 80% prediction intervals and observed values for total population sizes, while the lower two graphs show the scores of the two forecasts.

4. *The interest is in the* DSS *value for a scaled random variable* X/N *with scaled value* y/N *of* y (N *non-random and positive), written as* $DSS(y/N)$. *Then* $DSS(y/N) = 2\ln(\sigma/N) + [(\mu/N - y/N)/(\sigma/N)]^2 = DSS(y) - 2\ln(N)$. *For* N *we select expected population size* μ.

Figure III – **Total population size, Norway. Prediction intervals, observed values and scores**



Note: Prediction intervals and observed values in the upper panels, interval scores (*GRIS* and *GRISmod*), Dawid-Sebastiani (*DSS*), and variance-based (*VS*) scores in the lower panels. Prediction intervals, observed values, *GRIS*, *GRISmod*, and *VS* are in millions. Dawid-Sebastiani score is divided by ten.
Sources: See first paragraphs of Section 3.

Both forecasts underpredicted total population from around 2005 onwards. The most important explanation is that after the enlargement of the European Union, labour immigration to Norway from Baltic and East-European countries was much higher than expected. Note that at each forecast lead-time, the prediction intervals for StocProj are wider than the UPE-intervals. The modified interval score *GRISmod* rewards sharpness, and hence it is lower and thus better for UPE than for StocProj, although the difference is small; cf. the dotted lines. The modified interval score *GRISmod* and the variance-based score *VS* show the same trend: both forecasts become gradually worse for longer lead times. The dashed curves show the Dawid-Sebastiani score *DSS* divided by ten, so that we could plot it in the same graph as the other three scores. *DSS* starts at negative values in both cases, because the standard deviations σ of both population size forecasts are small (measured in millions) in the first few years. For instance, for StocProj in 1997, $\sigma = 0.0039$, which gives $\ln(\sigma^2) = -11.1162$. Since $((\mu\text{-}y)/\sigma)^2 = 0.0309$, *DSS* equals -11.0853, plotted as -1.1085 in Figure III. *DSS* increases steeply for UPE, because it does not reward sharpness anymore as soon as the observed value is more than one standard deviation away from the expected value ($|\mu - y| > \sigma$; cf. Section 3.1). This occurs in all years for which we have UPE-data, i.e. from 2004 onwards. For StocProj, the situation with too small standard deviation to reward sharpness does not occur until 2008, 12 years into the future. On the other hand, score functions *GRISmod* and *VS* do not punish "over-optimistic" forecasts (i.e. forecasts for which the variance of the predictive distribution is too small). Note that for StocProj, *DSS* stabilizes from around 2016, 20 years into the future.

For total population size predictions of the Netherlands, the UPE 80% prediction intervals reflect a sharper forecast than the 67% intervals of Statistics Netherlands' forecast; see Figure IV. In both cases, observed population size falls outside the intervals during a few years until 2011. Next, the observations are inside the intervals. The modified interval score for the UPE forecast is much better than that of Statistics

Netherlands' forecast. Interval scores miss the fact that observed values come closer to the centre of the intervals, because these scores do not include information about the mean, the median, or the mode of the predictive distribution. Judged by the Dawid-Sebastiani scores, the two forecasts are of equal quality. In both cases, *DSS* stabilizes from 2010 onwards. The reason is that the forecast error $|\mu - y|$ diminishes slowly over time, because observed population size approaches expected population size; this compensates the increase in the standard deviations of predicted population size in the two forecasts – see expression (3).

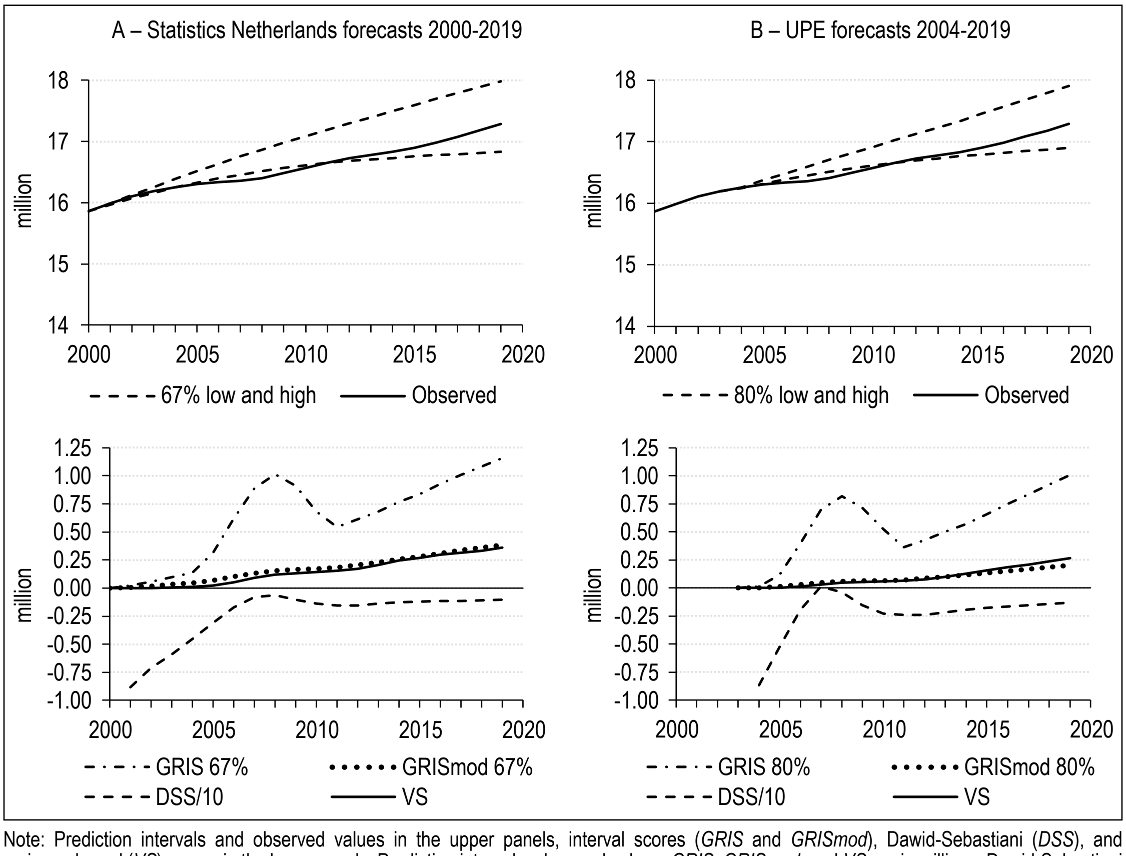
GRIS shows the same, rather irregular, time pattern as *DSS*, qualitatively speaking. This is very clear in Figure IV for the Netherlands, caused by the fact that observations first leave the intervals, but next return. Similar irregularities (but to a much smaller degree) are also visible in Figure III for Norway. In addition, *GRISmod* and *VS* develop very smoothly for the Netherlands, as we saw already for Norway.

Figure V gives UPE scores for total population size of Metropolitan France. A striking feature is that the forecast jump-off population in 2003 is almost 500,000 persons lower than the current estimate for population size that year. Data from Eurostat, available in 2004, provided the basis for the UPE-simulations. Observed numbers in Figure V stem from Insee (see https://www.insee.fr/en/statistiques/serie/000067670). Obviously, the 2003 population number as reported by Eurostat in 2004 has been revised in later years.

The jump-off error results in extremely bad values for the (non-modified) Gneiting-Raftery and the Dawid-Sebastiani score functions. What would these scores have been, in case the UPE-forecast would have started from the current estimate of total population size in 2003 (60.102 million) rather than the number that was actually used (59.635 million)? We can give an approximate[5] answer by lifting the 80% prediction interval up by 467,000 persons. Figure VI shows the results, with the same vertical scales as in Figure V. *DSS* improves dramatically, to 5.2 and 5.6 in 2005 and 2006, respectively

5. *Approximate, because we ignore the consequences for fertility and mortality of a higher jump-off population.*

Figure IV – **Total population size, Netherlands. Prediction intervals, observed values and scores**

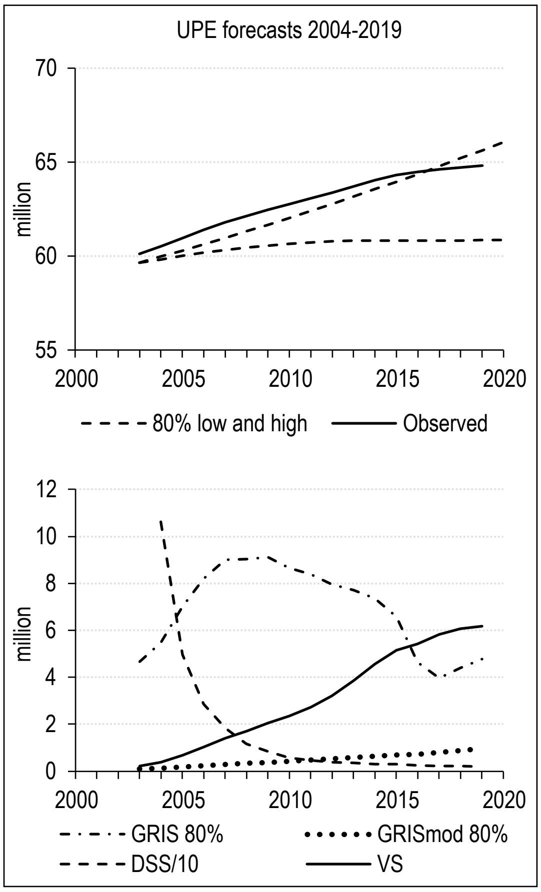


Note: Prediction intervals and observed values in the upper panels, interval scores (*GRIS* and *GRISmod*), Dawid-Sebastiani (*DSS*), and variance-based (*VS*) scores in the lower panels. Prediction intervals, observed values, *GRIS*, *GRISmod*, and *VS* are in millions. Dawid-Sebastiani score is divided by ten.
Sources: See first paragraphs of Section 3.

(instead of 49.6 and 28.6 for these years), while it stabilizes at a level around 1.6 / 1.7 after 2015 (rather than falling slowly to 2.0 in 2019). The interval scores and the variance-based score become slightly lower. These findings illustrate the importance of starting from the right jump-off population. At the same time, revision of population numbers occurs frequently, in particular in countries without a population register. In such cases, one should treat the jump-off population as stochastic, in addition to parameters for fertility, mortality and migration. Alho & Spencer (2005) give an example of random jump-off values for a probabilistic population forecast for Lithuania.

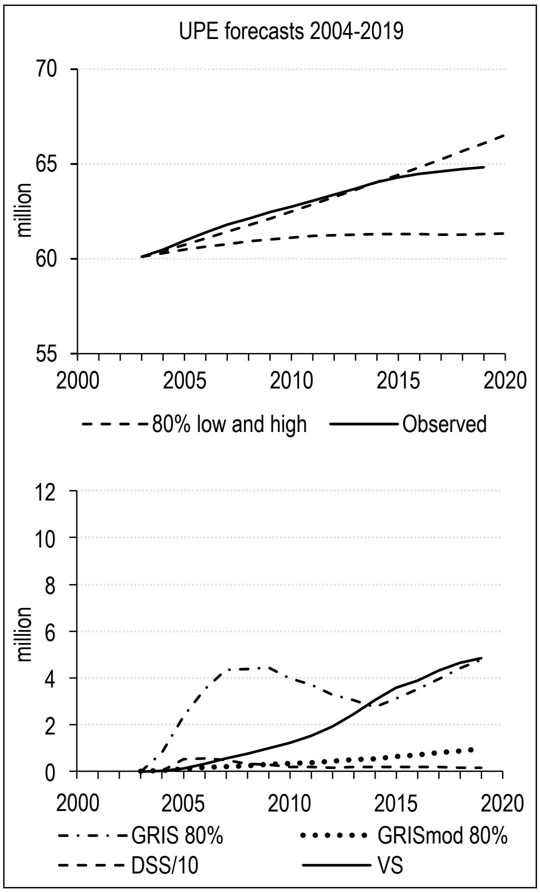
A common finding so far is that when we look further into the future, *GRISmod* and *VS* get worse over time, because prediction intervals become wider, and variances of predictive distributions increase. This, of course, reflects the fact that population forecasting is more difficult in the long-term than in the short-term. In contrast to *GRISmod* and *VS*, *DSS* stabilizes when forecast lead-times increase. The explanation lies in the definition of this particular scoring function. It is the sum of two terms: one term increases while the other one decreases when prediction variance goes up – see expression (3). Thus, one view is that *DSS* is not an appropriate measure for analysing how fast forecast quality deteriorates with increasing lead-time. However, a different view is that, exactly because *DSS* hardly changes over time, it controls for forecast lead-time. Still another possibility is to inspect the slopes in *GRISmod* and *VS*, since these two score functions increase quite smoothly with time. Further research into this issue, drawing upon data from many other forecasts (and controlling for different population sizes; see below) is clearly needed.

Figure V – **Total population size, metropolitan France. Prediction intervals, observed values and scores**



Note: Prediction intervals and observed values in the upper panel, interval scores (*GRIS* and *GRISmod*), Dawid-Sebastiani (*DSS*), and variance-based (*VS*) scores in the lower panel. Prediction intervals, observed values, *GRIS*, *GRISmod*, and *VS* are in millions. Dawid-Sebastiani score is divided by ten.
Sources: See first paragraphs of Section 3.

Figure VI – **Total population size, metropolitan France. Prediction intervals, observed values and scores. Adjusted jump-off population**



Note: Prediction intervals and observed values in the upper panel, interval scores (*GRIS* and *GRISmod*), Dawid-Sebastiani (*DSS*), and variance-based (*VS*) scores in the lower panel. Prediction intervals from UPE forecasts 2004-2019 are adjusted for jump-off error. Prediction intervals, observed values, *GRIS*, *GRISmod*, and *VS* are in millions. Dawid-Sebastiani score is divided by ten.
Sources: See first paragraphs of Section 3.

Table 1 – **Normalized interval scores, variance-based score, and Dawid-Sebastiani score**

	Norway		Netherlands		France[a]
	StocProj	UPE	CBS	UPE	UPE
year 2018					
GRIS/μ	0.564	0.513	0.062	0.053	0.069
GRISmod/μ	0.038	0.022	0.021	0.011	0.014
VS/μ^2 (x 1000)	17.552	6.569	1.108	0.781	1.154
DSS – 2ln(*μ*)	-1.525	2.073	-6.797	-7.149	-6.639
15 years ahead					
GRIS/μ	0.231	0.513	0.049	0.053	0.069
GRISmod/μ	0.021	0.022	0.016	0.011	0.014
VS/μ^2 (x 1000)	4.870	6.569	0.906	0.781	1.154
DSS – 2ln(*μ*)	-3.752	2.073	-6.903	-7.149	-6.639

[a] Adjusted for error in jump-off population.
Sources: See first paragraphs of Section 3.

As mentioned earlier, one explanation for the relatively bad scores for France is the fact that the score functions depend of population size. For a comparison across countries, normalized scores are useful. We normalized the scores the way explained in Section 2.4. Table 1 gives results for the five forecasts in 2018.

After normalization, the scores for the French forecast and the two Dutch forecasts in the year 2018 become very similar; see the upper panel of Table 1. In many cases, the scores for these two countries are one order of magnitude better than those for Norway. For many years, observed population size in France and the Netherlands fell within the prediction intervals (cf. upper panels of Figures IV and VI; the French intervals corrected for jump-off error). This contributes to the good scores for the two countries.

The two forecasts for Norway still receive bad scores because of the under-prediction of net immigration mentioned above. For the high StocProj scores in 2018 there is an additional reason: the jump-off year of this forecast is 1996, and hence the forecast lead time in 2018 is 22 years – much longer than the UPE lead time in 2018 (15 years). The lower panel of Table 1 shows the normalized scores for StocProj after a forecast duration of 15 years (in 2011). Compared to the scores for the other two countries after 15 years, the situation has improved quite much, but StocProj-scores are still much higher than those for StatNeth and for UPE in France and the Netherlands.

The final evaluation of total population size forecasts is by means of the *CRPS*. We computed this score function using 3,000 UPE-simulations for 2010. The *CRPS* depends of population size – see expression (6). To enhance comparison between the three countries, Table 2 gives normalized

Table 2 – **Normalized *CRPS* scores for total population size, UPE forecasts for 2010**

Norway	Netherlands	France
0.0249	0.0075	0.0492

Sources: See first paragraphs of Section 3.

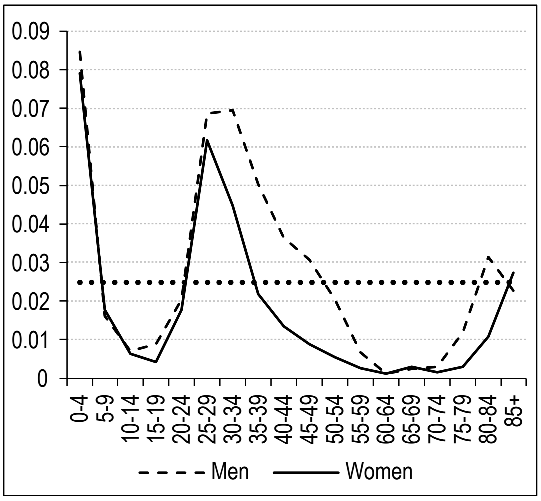
scores, defined as the *CRPS* divided by the mean of the 3,000 simulations. The results confirm the good quality of the UPE-forecast for the Netherlands that we found earlier.

3.2. Age and Sex Structures

Figures VII to IX plot normalized *CRPS* scores for simulated populations broken down by sex and five-year age group on 1 January 2010 according to the UPE forecasts. The horizontal dotted lines represent *CRPS* values for total population sizes from Table 2. The three graphs use the same vertical scale.

The age patterns of the scores differ strongly between the three countries. The findings for Norway in Figure VII are easy to interpret. High scores, i.e. low-quality forecasts, apply to young children, young adults, and the elderly. Scores are much better for ages 10-19 and 55-74. This age pattern reflects the under-prediction of immigration after 2005, already noted in Section 3.1. However, prediction errors for births and deaths may have contributed, too. Indeed, the age pattern of the scores is qualitatively similar to the pattern found for absolute errors in point forecasts of age and sex structures in industrialized countries (e.g. Keilman, 2009). This reflects the fact that births, migration flows, and deaths are difficult to predict. The lead-time of the UPE-forecasts is only seven years. At such a short horizon, fertility has no impact on the age group 10-19. International migration and mortality influence these age groups only very little. The same holds

Figure VII – **Normalized *CRPS* scores for population by age and sex, Norway, UPE forecast 2010**



Sources: See first paragraphs of Section 3.

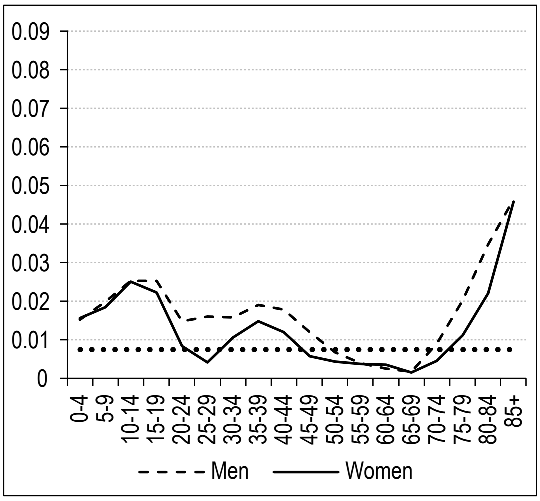
for age group 55-74. Clearly, had the evaluation taken place after a lead-time of twenty years or more, the normalized *CRPS* values for age groups 10-19 and 55-74 would have been much worse. Finally, note that the scores for men in ages 19-54 and 75+ are somewhat higher than those of women in these age groups. The reason is that men are more prone to migrate (19-54) or to die (75+) than women.

Whereas the Norwegian score agrees with what one might expect, the scores for the other two countries are more difficult to interpret. Normalized scores indicate that the Dutch forecast is of better quality than the other two, except for old ages. The French scores tend to decline with age. This pattern suggests that fertility was more difficult to predict accurately, than international migration or mortality. One may also think of several other explanations. First, the revision of the population numbers discussed above may have been stronger in some age groups than in others. We found (numbers not shown here) that revised numbers for men and women by five-year age group are approximately one percent higher than those used in UPE. However, there are a few exceptions. Revisions were less than half a percent in age groups 0-4 and 80+, while for men aged 20-24 the revised number was one percent lower than the number used in UPE. This pattern, caused by revisions between 2003 and 2010, is not reflected in Figure IX. A second explanation is that under- or over-prediction of net migration flows to France during the years 2003-2009 may also differ across age groups. Finally, our empirical data on age-sex distributions as of 2010 include the effects of so-called administrative corrections, this expression covering both corrections for errors in registrations and statistical adjustments. Such corrections are necessary in case registration of births and deaths is incomplete. For register countries (Norway and the Netherlands), errors in registered immigration and emigration are included as well in the administrative corrections. For Norway, the effect of these corrections is likely small. It is much larger for the Netherlands and France: for instance, data from Statistics Netherlands and Insee show that total net migration for the years 2003-2009 without corrections amounts to 214,000 and 601,000 persons respectively. But Eurostat provides net migration data including corrections. Using those data, we find that the totals for net migration during 2003-2009 are very different, namely 17,000 (the Netherlands) and 884,000 (France).[6] Because of the lack of reliable data distinguishing net migration and administrative corrections broken down by age for the Netherlands and France, we have not analysed this issue further. Note also that the UPE-forecasts do not include a separate variable for administrative corrections (as is common practice for population forecasting).

The general conclusion from this evaluation is that the UPE-forecast of the Dutch population pyramid for 2010, as measured by the normalized *CRPS* score, is better than the UPE-forecasts of Norway and France, except for the oldest-old. The age pattern for the Norwegian *CRPS* score is similar to that of absolute errors in point

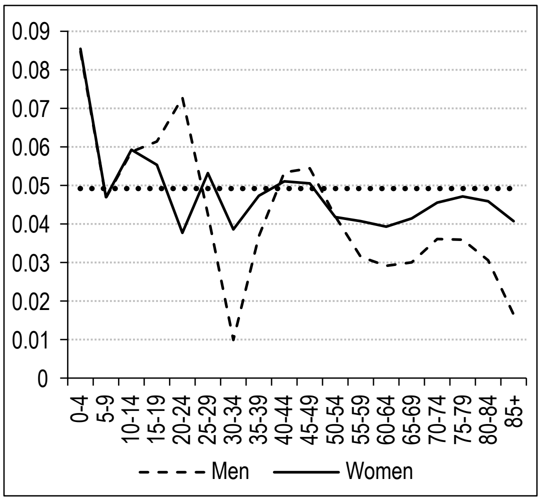
6. For Norway, the numbers are 188,300 (without corrections) and 187,800 (with corrections). For France, Insee provides separate estimates of net migration and adjustments. These are not detailed in Eurostat data.

Figure VIII – **Normalized *CRPS* scores for population by age and sex, Netherlands, UPE forecast 2010**



Sources: See first paragraphs of Section 3.

Figure IX – **Normalized *CRPS* scores for population by age and sex, Metropolitan France, UPE forecast 2010**



Sources: See first paragraphs of Section 3.

forecasts. It is difficult to indicate why the age patterns differ strongly between the three countries, due to data problems for international migration in particular.

* *
*

The purpose of this paper is to demonstrate how a probabilistic population forecast can be evaluated, when observations for the predicted variables become available. Statisticians have developed various scoring rules for that purpose, but there are hardly any applications in population forecasting literature. A scoring rule measures the distance between the probability distribution of the predicted variable, and the actual outcome. A score as such has no intrinsic meaning – we can only interpret it by comparing it to the score of another forecast. We have used scoring rules that reward accuracy (the outcome is close to the expected value of the prediction) and sharpness (the predictive distribution has low variance, which makes it difficult to hit the target). One may argue that accuracy is more important than sharpness: sharpness ought to be rewarded only when the outcome is not too far away from the central tendency of the predictive distribution. We discussed the notion of "too far away".

A forecaster can make the probabilistic forecast available to the user in three different ways. The first is by publishing a prediction interval for the variable of interest. Coverage probabilities of 67 and 80% are rather common. Some population forecasters present 95% prediction intervals. We do not recommend this practice, because 95% intervals are very wide as they stretch to quantiles where extreme events start to happen. The second method is to give the user access to a database that contains sample paths for the stochastically simulated development in population size and other forecast results. Sometimes, only the first moment (expectation) and the second moment (variance) of the prediction interval are available. We presented scoring rules that one may use for either type of forecast results. The scoring rules are negatively oriented: a lower score implies a better forecast.

We have evaluated probabilistic population forecasts for France, the Netherlands, and Norway. For all three countries, we have used results from the UPE-project. Since many scoring rules apply the same scale as population size, we proposed using normalized scoring rules when the interest is in comparing forecasts for different countries. We inspected prediction intervals for population size in the period 2004-2019 and 3,000 sample paths for population pyramids for the year 2010. For the Netherlands and for Norway, we compared the UPE-results with findings from the official probabilistic population forecast by Statistics Netherlands (2001-2019) and from a probabilistic forecast for Norway (1997-2019). All forecasts were computed using the cohort-component method and stochastically varying parameters for fertility, mortality and migration, and a deterministic jump-off population.

Our evaluations show that the UPE-forecasts for the Netherlands and for Norway performed better than the other forecasts for these two countries, because the UPE-predictions were relatively sharp, with narrow prediction intervals. The UPE-forecast for France started from a jump-off population in 2003 that was estimated at 60.1 million persons at the time the forecast was computed. This number is almost 500,000 persons higher than the current estimate of the population in 2003 (59.6 million). The error in the jump-off population caused a bad score for the French forecast. To revise population statistics for inter-census years when data from a new population census become available, is common practice. In case one cannot be certain about the size and structure of a population during an inter-censal period, the correct approach is to treat the jump-off population of the forecast as stochastic.

We evaluated the 3,000 UPE-simulations of the age and sex composition predicted for the year 2010. When normalized for population numbers in each age-sex category, the predictions for the Netherlands received the best scores, except for the oldest old. The age pattern for the Norwegian score reflects the under-prediction of immigration after the enlargement of the European Union in 2005. However, prediction errors for fertility and mortality may have played a role as well. The age-specific scores for France are difficult to interpret. They do not reflect the age pattern of the revision of the population data for 2003 mentioned above. Over- or under-prediction of fertility, mortality and migration may have played a role. In the cohort-component model, the age- and sex-composition of the population of 2010 is a complicated non-linear function of model parameters for mortality, fertility, and migration prior to 2010. Therefore, one cannot identify the contribution of these three components of change to the scores.

In addition to the issue of data revision, we were also confronted with the problem of "administrative corrections". This is a notion that statistical agencies sometimes use as a distinct component of change of population size and structure. When there are errors in the registration of births, deaths, and migrations, administrative corrections and statistical adjustments are necessary to obtain a correct set of bookkeeping statistics for population. Empirical population numbers for the Netherlands and France are strongly influenced by these corrections.

There is a rich literature that evaluates probability forecasts and that discusses a large number of scoring rules. Many apply to predictive distributions of a discrete random variable, and are of little interest for evaluating demographic forecasts. In case we limit ourselves to scoring rules for continuous random variables, the literature still proposes many scoring rules, of which we selected just a few. As we have shown in Sections 2 and 3, these scoring rules are very different, giving different weight to distance or to sharpness. Some rules give a bad score as soon as observed numbers fall outside the prediction interval. Others develop more smoothly when the observation is further and further away from the central tendency and from the interval bounds. Further work applied to scoring rules for probabilistic demographic forecasts is necessary, hopefully leading to guidelines for the selection of such rules in various situations.

Scoring rules are useful in *ex-post facto* evaluations of two or more probabilistic forecasts. Once we have concluded that, judged by a number of score functions, one forecast was better than another one, we have to ask ourselves why this was the case. To answer that question, one needs to analyse very carefully the many steps in the production process of the two probabilistic forecasts. This poses a new challenge, in particular when different scholars or different agencies computed the two forecasts. ☐

BIBLIOGRAPHY

Alders, M. & De Beer, J. (1998). Kansverdeling van de bevolkingsprognose ("Probability distribution of the population forecast"). *Maandstatistiek van de Bevolking*, 46, 8–11.

Alexopoulos, A., Dellaportas, P. & Forster, J.J. (2018). Bayesian forecasting of mortality rates by using latent Gaussian models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(2), 689–711. https://doi.org/10.1111/rssa.12422

Alho, J. & Nikander, T. (2004). Uncertain population of Europe: Summary results from a stochastic forecast. http://www.stat.fi/tup/euupe/del12.pdf

Alho, J. & Spencer, B. (2005). *Statistical Demography and Forecasting*. New York: Springer.

Askanazi, R., Diebold, F. X., Schorfheide, F. & Shin, M. (2018). On the comparison of interval forecasts. *Journal of Time Series Analysis*, 39(6), 953–965. https://doi.org/10.1111/jtsa.12426

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. & Buisson, G. (2016). Projections de population à l'horizon 2070 : Deux fois plus de personnes de 75 ans ou plus qu'en 2013. *Insee Première* N°1619. https://www.insee.fr/fr/statistiques/fichier/version-html/2496228/ip1619.pdf

Brier, G. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1), 1–3. https://doi.org/10.1175/1520-0493(1950)078%3C0001:VOFEIT%3E2.0.CO;2

Costemalle, V. (2020). Projections probabilistes bayésiennes de population pour la France. *Economie et Statistique / Economics and Statistics*, this issue.

Gneiting, T. & Raftery, A. (2007). Strictly proper scoring rules, prediction and estimation. *Journal of the American Statistical Association*, 102(477), 359–378. https://doi.org/10.1198/016214506000001437

Gneiting, T. & Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Applications*, 1, 125–151. https://doi.org/10.1146/annurev-statistics-062713-085831

ISTAT – Istituto Nazionale di Statistica (2018). Il futuro demografico del paese: Previsioni regionali della popolazione residente al 2065 (base 1.1.2017). *Report Statistische* 3 maggio 2018. Roma: ISTAT.

Jordan, A., Krüger, F. & Lerch, S. (2019). Evaluating Probabilistic Forecasts with scoringRules. *Journal of Statistical Software*, 90(12). https://doi.org/10.18637/jss.v090.i12.

Keilman, N. (1990). *Uncertainty in national population forecasting: Issues, backgrounds, analyses, recommendations*. Amsterdam and Rockland, MA: Swets and Zeitlinger Publishers.

Keilman, N. (2009). Erroneous population forecasts. *In*: P. Festy & J.-P. Sardon (Eds.) *Profession démographe - Hommage à Gérard Calot*, pp. 237–254. Paris: INED.

Keilman, N., Pham, D. Q. & Hetland, A. (2002). Why population forecasts should be probabilistic - illustrated by the case of Norway. *Demographic Research*, 6-15, 409–454. https://doi.org/10.4054/DemRes.2002.6.15

Keyfitz, N. (1981). The limits of population forecasting. *Population and Development Review*, 8(44), 579–593. https://doi.org/10.2307/1972799

Matheson, J. E. & Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Management Science*, 22, 1087–1096. http://dx.doi.org/10.1287/mnsc.22.10.1087

Murphy, A. (1970). The ranked probability score and the probability score: A comparison. *Monthly Weather Review*, 98(12), 917–924. https://doi.org/10.1175/1520-0493(1970)098%3C0917:TRPSAT%3E2.3.CO;2

NRC – National Research Council (2000). *Beyond six billion: Forecasting the world's population. Panel on Population Projections*. John Bongaarts and Rudolfo Bulatao (eds). Washington DC: National Academy Press.

Raftery, A., Li, N., Ševčíková, H., Gerland, P. & Heilig, G. (2012). Bayesian probabilistic population projections for all countries. *PNAS - Proceedings of the National Academy of Sciences*, 109(35), 13915–13921. https://doi.org/10.1073/pnas.1211452109

Shang, H. L. (2015). Statistically tested comparisons of the accuracy of forecasting methods for age-specific and sex-specific mortality and life expectancy. *Population Studies*, 69(3), 317–335. https://doi.org/10.1080/00324728.2015.1074268

Shang, H. L., Smith, P., Bijak, J. & Wisniowski, A. (2016). A multilevel functional data method for forecasting population, with an application to the United Kingdom. *International Journal of Forecasting*, 32, 629–649. https://doi.org/10.1016/j.ijforecast.2015.10.002

Shang, H. L. & Hyndman, R. (2017). Grouped functional time series forecasting: An application to age-specific mortality rates. *Journal of Computational and Graphical Statistics*, 26(2), 330–343. https://doi.org/10.1080/10618600.2016.1237877

Smith, S., Tayman, J. & Swanson, D. (2001). *State and Local Population Projections: Methodology and Analysis*. New York: Kluwer Academic/Plenum Publishers.

Staël von Holstein, C.-A. (1970). A family of strictly proper scoring rules which are sensitive to distance. *Journal of Applied Meteorology*, 9, 360–364. https://www.jstor.org/stable/26174966

Statistics Netherlands (2001). Maandstatistiek van de Bevolking N° 49 (januari), pp. 63–70.

Statistics New Zealand (2011). National Population Projections: 2011(base) – 2061. Bulletin published 19 July 2012. http://archive.stats.govt.nz/browse_for_stats/population/estimates_and_projections/NationalPopulationProjections_HOTP2011.aspx (accessed on 21 March 2019).

Is the Ageing of the French Population Unavoidable?

Nathalie Blanpain*

Abstract – A population projection is not a certain prediction, but rather an estimate of what the future evolution of the population might be under certain assumptions about changes in mortality, fertility and migration, around a central scenario that suggests a continuation of recent demographic trends. This article looks at the assumptions made for the population projections conducted for France in 2016. It first reviews the approach used by Insee to establish them, and then examines the more or less certain nature of the main results. The ageing process observed for more than a century is expected to continue; however, if an indicator based on "prospective age" is used, the population would not age. The evolution of the population as a whole is uncertain. In 2070, the size of the population of the 28-member European Union would be close to that of 2019. The improvement in life expectancy combined with a positive migratory balance would compensate for a fertility level that does not allow for the renewal of generations.

JEL Classification: J11, N34
Keywords: population projections, ageing, fertility, mortality, life expectancy, prospective age, net migration, France, European Union

**Insee (nathalie.blanpain@insee.fr)*

Received October 2019, accepted April 2020.
Translated from: "Le vieillissement de la population française est-il inéluctable ?"
Citation: Blanpain, N. (2020). Is the Ageing of the French Population Unavoidable? *Economie et Statistique / Economics and Statistics*, 520-521, 65–85. https://doi.org/10.24187/ecostat.2020.520d.2029

Population projections provide population estimates over various time periods, based on different assumptions. These are therefore not certain forecasts and some events may lead to significant differences between actual data and the projected data. One extreme example is the projection made by Alfred Sauvy in 1936, presented by Hubert *et al.* (1937), in a chapter entitled "*La dépopulation à craindre et les remèdes à lui opposer*" [Depopulation to be feared and the remedies to counter it]. If the demographic trends from that time had continued, France would have had around 29.6 million inhabitants in 1985 (see Appendix). However, there have been 25.6 million more than that. The projection assumed that fertility would continue to decline at the same rate as in the 1930-1935[1] period and obviously did not anticipate the post-war baby boom. Furthermore, it assumed that mortality would continue to decline at the same rate as in the 1925-1935 period. The authors even thought that this continuation of the decline in mortality was optimistic: "the projected number of deaths in 1985 may seem unrealistically low, as it corresponds to a 65% reduction in age-specific mortality for both men and women under the age of 50" (Hubert *et al.*, 1937, p. 217). However, the projection proved to be pessimistic, since mortality fell at a greater average annual rate between 1935 and 1985 than over the 1925-1935 reference period. According to the 1937 projection, the death toll in 1985 would have been 556,000, giving a mortality rate of 1.9%, almost two times higher than the rate actually recorded in 1985 (1.0%). Similarly, net migration was assumed to be zero. The authors indicate that "if population growth continues to slow down more and more in Europe, the source from which we have drawn our migrants will quickly dry up". Finally, net migration was clearly positive every year between 1946 and 1985.

This historic example illustrates the importance of the assumptions made in making population projections. These projections are very important for informing public decisions, such as those concerning, for example, the balance of the pension system, the number of educational institutions, early childcare facilities, etc. A demographic projection typically refers to the population, broken down by sex and age. Additional modelling can enrich the projection, in accordance with other variables of interest, such as region of residence (Desrivierre, 2017), professional activity (Koubi, 2017), state of health and level of dependency (Roussel, 2017), for example.

Two major approaches are possible for estimating the future population: deterministic and probabilistic. The deterministic approach makes it possible to estimate "what would happen" under a set of assumptions that define a scenario; this is the approach used for the population projections published by Insee in 2016. Several sets of assumptions make it possible to develop several scenarios. The most robust results are those that are obtained in all scenarios, while the weakest are those that vary greatly depending on the scenario. Assumptions can be developed based on extrapolation of past trends, the establishment of long-term trends (based, in particular, on expert opinion) or a structural model that explains population change using exogenous variables, and often a combination of these elements (Costemalle, this issue).

Probabilistic approaches quantify uncertainty over "what would happen" with a given probability. In this case, a large number of projections must be made to calculate a confidence interval. Here, the sets of assumptions are based on the modelling of fertility, mortality and migration. For France, the projections resulting from the two approaches are not very far apart: thus, the population size in metropolitan France in 2050 obtained using the central scenario of the deterministic approach differs from that obtained in the median scenario of the probabilistic approach by only 2% (Costemalle, this issue).

Whether the deterministic or probabilistic approach is used, the component method is generally applied. This involves "ageing" the last-known age pyramid from year to year, with the aim of determining the age pyramid for a certain number of years. The Swedish statistician Sven Wicksell was one of the first to use this method to estimate the evolution of the Swedish population in 1926 (Wicksell, 1926; Wattelar, 2004). Only a few events can change a country's population upwards or downwards: births, deaths and migration. The assumptions therefore concern future developments in fertility, mortality and net migration. The population is then changed by sex and age, by adding births by sex, subtracting deaths by sex and age and adding net migration by sex and age.

This article primarily focuses on the assumptions of the population projections established for France in 2016. The first section reviews the approach adopted by Insee to establish those projections. The second section is devoted to the main results, distinguishing between those that

1. The fertility rate fell from 2.3 to 2.1 over the period.

are relatively robust and those that are weakest. Finally, the third section compares France's situation with that of its European Union (EU) neighbours using projections published by Eurostat in 2019.

1. The Assumptions of the 2016 Population Projections for France

To develop the assumptions, Insee called upon both national experts, researchers and representatives of various institutions using the projections or specialists in certain fields,[2] and international experts, most of whom are responsible for population projections in their countries. Twenty-five of them responded to a questionnaire on the evolution of mortality, fertility, migration, the projection horizon and the method to be used. The responses, which are summarised here, are detailed in Blanpain & Buisson (2016a). Population projections are revised approximately every 5 years in France.

A projection horizon of 2070 was appropriate for most of the experts who gave an opinion on this subject. Two experts would have preferred a longer projection horizon and three would have preferred a shorter one. The projection horizon of 2070 was therefore used.

Most of the experts agreed on the complementarity of the deterministic method and the probabilistic method. The deterministic method was chosen because it allows for easier communication to a non-specialist audience. It also makes it easier to make derived projections (e.g. active population projections).

This projection is based on the component method. It consists in estimating the population for the following year (year n+1) based on the starting population (year n), then adding births and net migration (immigration – emigration) and subtracting deaths, then repeating the operation year after year:

$$Pop_{n+1} = Pop_n + Birth_n - Death_n + Net\,Migration_n$$

In France, population estimates and statistics from the civil status registry make it possible to estimate age-specific fertility rates in previous years and to establish the history of mortality rates, i.e. the probability of dying within the year by sex and age. Net migration rates by sex and age are established by the difference in successive populations and the natural balance (births – deaths).

Most of the experts approved the choice to use an odd number of assumptions, allowing a central scenario to be defined. Three assumptions (central, low and high) were made for each of the components, mortality, fertility and migration. The central assumption is generally that of a continuation of recent trends. The low assumption uses a slower evolution than in the past and the high assumption uses a faster one. Projections based on the continuation of trends, as in this case, are unable to predict trend reversals by definition. The analysis of the differences between the evolutions observed and earlier projections (Blanchet & Le Gallo, 2014) calls for caution, which leads to the use of several scenarios to analyse the sensitivity of results to different assumptions.

One scenario is based on one assumption for fertility, one for mortality and one for net migration. The combination of the three assumptions (central, low and high) for each component results in twenty-seven scenarios. Of these scenarios, the central scenario combines the central assumptions of the three components. Six scenarios illustrate what would happen if only one of the assumptions was changed compared to the central scenario: the low and high life expectancy scenarios, the low and high fertility scenarios and the low and high migration scenarios. In addition, four alternative scenarios combine the assumptions leading to a low, high, young or elderly population. For example, the elderly population scenario combines an assumption of high life expectancy, low fertility and low migration.

Finally, three other scenarios were also constructed, making it possible to estimate what would happen if France's fertility rate was the same as the European fertility rate in 2015, if life expectancy remained at its 2014 level or if net migration were zero.[3]

1.1. Mortality

The central assumption assumes that mortality will continue to fall at the same rate as in the past until the projection horizon. This therefore requires the definition of a reference period for said past. The reference period chosen here is 1995-2014. This includes the year 2003, when there was a heat wave and mortality increased particularly at high ages, as well as the following

2. Including: The Conseil d'orientation des retraites [Pension Advisory Board] (COR), the Direction de la recherche, des études, de l'évaluation et des statistiques [Directorate of Research, Studies, Evaluation and Statistic] (Drees), the Institut national d'études démographiques [National Institute for Demographic Studies] (INED), the Institut Paris région [Paris Region Institute] and the Institut national de la santé et de la recherche médicale [National Institute of Health and Medical Research] (INSERM).
3. The assumptions and results for all scenarios have been published (Blanpain & Buisson, 2016b).

year 2004, when life expectancy rebounded exceptionally, by +11 months for both men and women (Papon, 2019). In the end, the heat wave episode paradoxically had a rather positive long-term effect on the evolution of life expectancy thanks to preventive measures aimed at the elderly in particular (Pison, 2007). The reference period is quite long, twenty years from 1995 to 2014, so as to smooth out the impact of 2003-2004. However, the most recent trends are somewhat different: in particular, life expectancy is stagnating or increasing less quickly in some European countries, including France. According to Eurostat, life expectancy in the 28-member EU is 81.0 years in 2018, which is the same level as in 2014 (80.9 years). In France, between 2014 and 2019, life expectancy rose by only 0.2 years for women and 0.5 years for men (Beaumel & Papon, 2020). Indeed, three of the five years from 2014 to 2018 were marked by a relatively deadly flu epidemic (Équipes de surveillance de la grippe, 2018). However, the slower progress in life expectancy may also be a sign that the benefits of the "cardiovascular revolution" are coming to an end (Pison, 2019). Furthermore, among women, mortality linked to cancer has stopped falling in recent years, particularly due to the rise in smoking in the 1950s to 1980s among those aged 50 or older today (Pison, 2019). The reference period chosen therefore leads to a slightly more optimistic projection than if the latter data had been known. At the time the assumptions were constructed, this stagnation was not anticipated, or at least not as a sustainable phenomenon to be included in the central long-term population projection. The question of the sustainability of the slowdown in improvements to life expectancy will arise in the next projection exercise.

The selection of assumptions for a projection is also partly explained by the lessons learned from past projections, in particular from the errors made at that time. Thus, the projections made in the 1970s and 1980s in France assumed that life expectancy would reach a ceiling in the more or less long term, believing that it was approaching a biological limit. However, that level proved to be far below the values observed subsequently (Blanchet & Le Gallo, 2014). For example, the 1979 projection resulted in a life expectancy of 78 years for women and 70 years for men in 2015, which is 7 years and 9 years less, respectively, than was ultimately observed. Starting in the 1990s, therefore, the projects adopted the approach of extrapolating past mortality trends without capping them, leading to results much closer to the observed data.[4] The 2016 projection is therefore based on a continuation of the mortality trends without a cap.

However, a novelty has been introduced, following the recommendations of one of the experts: the projection of mortality rates according to past trends has been amended to take into account a generational effect. Indeed, while age-specific mortality generally decreases from generation to generation, it stagnates in adulthood for generations born at the end of the Second World War or just after, for both men and women. For example, this stagnation is visible at age 50 for women (Figure I). At that age, the probability of dying within one year was 2.5 per 1,000 for women born in 1941, which is virtually identical to that for women born in 1956 (2.4 per 1,000, or -2%), while it fell for the previous generations born between 1931 and 1941 (-21%) and for later generations born from 1956 to 1966 (-21%). This plateau is observable for most adult ages, indicating a generational effect not related to the time period. One way to summarise this generational effect is to observe the probability of dying between two given ages (Figure II). For example, among women who have reached the age of 18, the probability of dying between the ages of 18 and 54 falls fairly little between the generations born from 1941 to 1956 (-9% in 15 years) and rapidly between the previous generations born from 1931 to 1941 (-22% in 10 years) and the following generations, born from 1956 to 1965 (-18% in 9 years).

This specific evolution is taken into account in the projections. The generational effect that is visible up to the age of 70[5] is thus assumed to continue until the end of the life of the so-called "plateau" generations, born between 1941 and 1956 for women and between 1941 and 1953 for men. In concrete terms, for the central assumption, the average annual rate of change in mortality at age 59[6] is calculated between the 1941 and 1956 generations for women (between 1941 and 1953 for men) and the same rate is applied to the following ages (see Figure I).

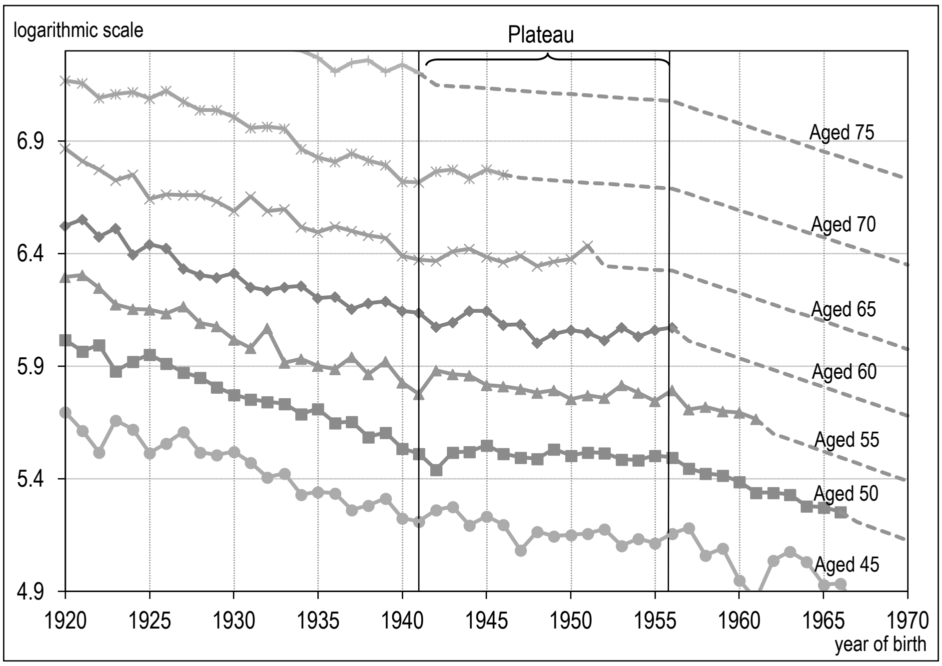
The annual rate of change in mortality rates around the age of 50 for the generations born from 1956 onwards is yet to be determined. Indeed, applying the rates of change in the mortality rates observed during the reference period would

4. Thus, for women, the projections for 1995, 2003, 2006 and 2010 all led to a life expectancy close to 85.5 years in 2015, which is a difference of less than one year compared to the observed situation. For men, the 1995 projection was somewhat pessimistic (a difference of two years) and the projections for 2003, 2006 and 2010 are close to the observed reality (a difference of less than one year).

5. At this age, only the start of the plateau is observed.

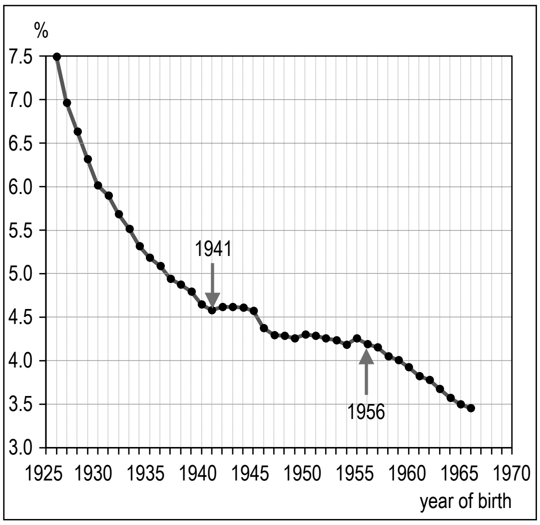
6. Age reached in 2015 by the generation born in 1956.

Figure I – **Female mortality rate by age and year of birth**



Reading Note: The probability of dying at age 50 for women born in 1966 is 1.9 per 1,000. This is calculated as follows: exp(5.3)/100.
Sources and coverage: Insee, population estimates and civil status registry statistics from 1965 to 2016; Insee, central population projection scenario from 2017 onwards. Metropolitan France for years up to 1990, France excluding Mayotte from 1991 to 2013, France from 2014 onwards.

Figure II – **Probability of dying for women aged 18 to 54 by year of birth**



Reading Note: Among women born in 1966 and alive at age 18, the probability of dying between the ages of 18 and 54 is 3.5%.
Sources and coverage: Insee, population estimates and civil status registry statistics. Women alive at age 18, Metropolitan France for years up to 1990, France excluding Mayotte from 1991 to 2013, France from 2014 onwards.

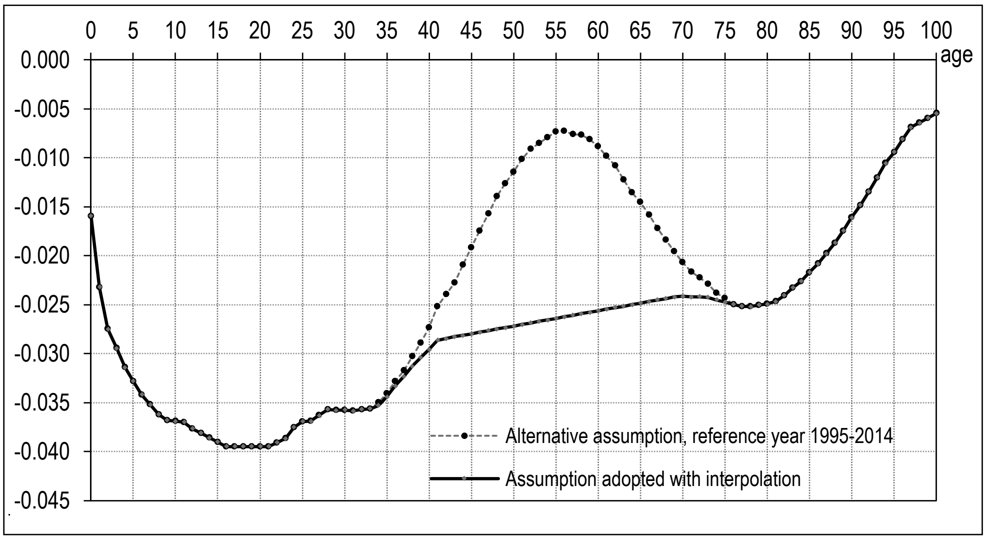
slow the decrease significantly. For example, to calculate the evolution of mortality rates at age 50 that will be experienced by the generations born from 1970 to 2020, the 1995-2014 reference period concerns the generations born from 1945 to 1964 who turned 50 during that period. This largely includes the "plateau" generations, for whom the decrease has slowed, while there is no reason to assume that this slowdown will affect later generations. The assumption used is that mortality resumes its downward trend for these generations. Thus, mortality at age 50 is declining at a steady rate, as was already the case before the plateau generations reached that age. The rate of decline is determined by interpolation between two ages (Figure III).

An alternative assumption, simply continuing past trends without taking the generational effect into account, has been tested. The assumption used and the alternative assumption lead to virtually the same life expectancy at birth in 2070 (Blanpain & Buisson, 2016a). Taking into account the generational effect leads to two compensatory effects: a slowdown in the decline in mortality for the generations born at the end of the Second World War or just after, and an acceleration in the decline in mortality at the age of around 55 for later generations. The evolution of life expectancy at age 60 is a little slower when using the chosen method (taking into account the generational effect) compared to the alternative method, particularly at the beginning of the period. For example, in 2037, the difference is -0.6 years for men and -0.8 years for women.

In summary, the assumption chosen as the central assumption is as follows:

- At each age, mortality continues to fall at the same rate as in the period 1995-2014, unless

Figure III – **Annual evolution of the female mortality rate logarithm by age**



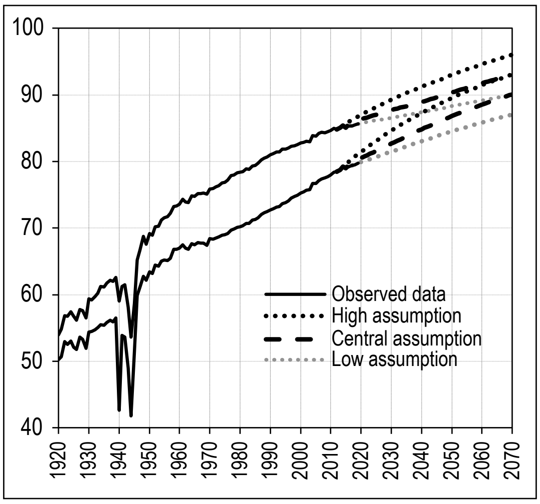
Reading Note: The annual decrease in the mortality rate logarithm for women aged 50 is -0.011 over the 1995-2014 period, on average.
Sources and coverage: Insee, population estimates and civil status registry statistics. Metropolitan France for years up to 1993, France excluding Mayotte from 1994 to 2013, France from 2014 onwards.

these years include the 1941-1956 generations for women (1941-1953 for men).

- If these years at least partially include these generations, the decline is calculated by interpolation.

- For the 1941-1956 generations, for women (1941-1953 for men), mortality is virtually stable at each age and the central assumption is that it will remain so.

Figure IV – **Life expectancy at birth according to different assumptions**



Reading Note: In France, female life expectancy at birth is 85.6 years in 2019.
Sources and coverage: Vallin & Meslé, French mortality tables for years until 1945; Insee, population estimates and civil status registry statistics from 1946 to 2019; Insee, population projections from 2013 to 2070. Metropolitan France for years up to 1993, France excluding Mayotte from 1994 to 2013, France from 2014 onwards.

The central assumption results in life expectancy at birth of 90 years for men and 93 years for women in 2070, which is an increase of 10.4 years for men and 7.4 years for women since 2019 (Figure IV). By way of comparison, between 1968 and 2019, a period of the same length (51 years), life expectancy for men increased slightly faster (11.9 years), while it increased significantly faster for women (10.4 years). The differences in life expectancy between men and women have reduced since the mid-1990s. Since then, male mortality has fallen more rapidly than female mortality, thanks in particular to the reduction in violent deaths and deaths due to cancer or AIDS (Meslé, 2006). According to the central assumption, life expectancy for men will become even closer to that for women, with the difference being just 3 years in 2070, compared to 6 years in 2019. Consequently, the rebalancing between men and women at older ages should continue. In 2070, 39% of people aged 95 would be men, compared to only 23% in 2020.

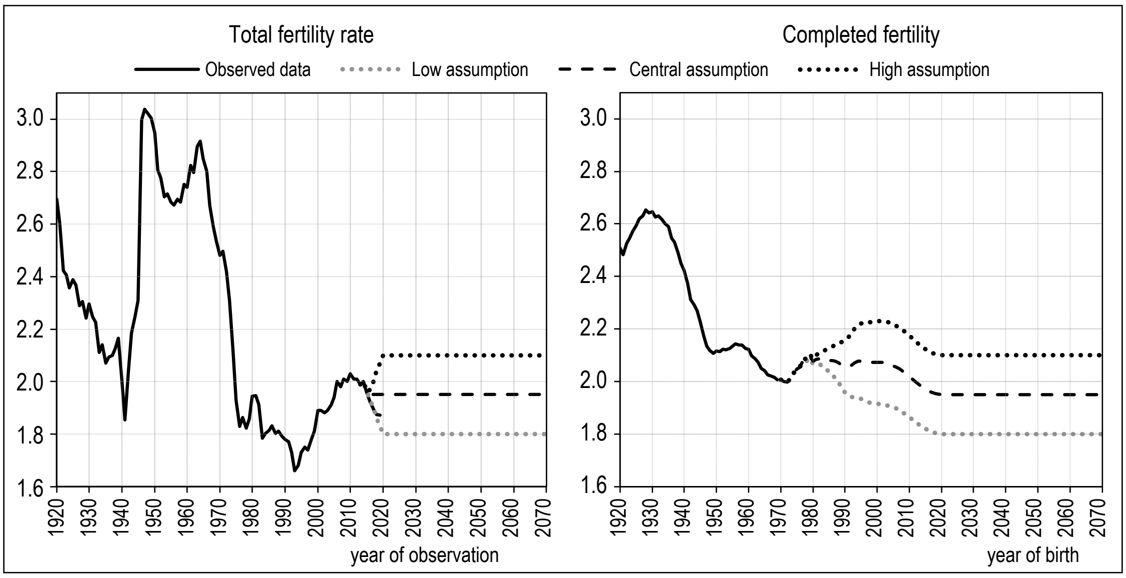
Low and high assumptions are considered for each of the components. The low assumption for mortality assumes that mortality rates will decrease at a lower rate than in the past, while the high assumption assumes that it will fall at a faster rate. The age-specific mortality rates are multiplied by the same coefficient so that the low and high assumptions lead to a life expectancy of plus or minus 3 years compared to the central assumption in 2070, i.e. between 87 and 93 years for men and between 90 and 96 years for women (Figure IV). An assumption for life expectancy that is constant and at the 2014 level, i.e. 79 years for men and 85 years for women, completes these three assumptions. In 2019, the life expectancy of men is 79.7 years and that of women 85.6 years in France, which is the level of the low assumption, given the recent slowdown in life expectancy improvements (Papon & Beaumel, 2020).

1.2. Fertility

As with mortality, the central assumption assumes that the age-specific fertility rates will evolve at the same rate as in the past. However, despite steady medical progress over recent decades, the experts agree that the average age at childbirth cannot increase indefinitely, as fertility declines with age. As a result, the trends are not continued to the projection horizon: fertility rates are stabilised once an average age at childbirth considered as a ceiling is reached. The experts were therefore questioned both on the level of fertility, as measured by the total fertility rate or by completed fertility, and on the evolution of the average age at childbirth. The total fertility rate reflects the average number of children a woman would bear if she knew the fertility conditions in a given year throughout her entire fertile life. It measures women's fertility level at a given moment. Completed fertility is the average number of children born by women of the same generation. It can therefore be calculated when they reach the end of their fertile life, i.e. at the age of 50.

Breaking a historical downward trend, the total fertility rate rose sharply from 1941, marking the beginning of the baby boom (Figure V). This ended in the 1970s: in 1976, the total fertility rate was only 1.83 children per woman, compared to around 2.48 still in 1970, for example. The total fertility rate then remained in a range from 1.8 to 2.0, except around 1993 when it was low (1.66) due to a temporary postponement of the birth schedule for generations born in the early 1970s, apparently linked to poor economic conditions (Pison, 2017). Earlier projections therefore used a central assumption within this range: 1.8 children per woman on three occasions in 1986, 1995 and 2003, then 1.9 children in 2006 and 1.95 in 2010 (Blanchet & Le Gallo, 2014). Assumptions of 1.90 and 1.95 children per woman in the last three projections (2006, 2010 and 2016) confirm and continue the high fertility of the years 2004-2014. Since then, fertility has fallen slightly, but this development is not (yet) taken into account in the projections. Completed fertility decreased overall from the generation born in 1930, of childbearing age throughout the baby boom period (2.6 children on average) to the generation born in 1970 (2.0 children, Figure V). It should increase to 2.1 children for the generation born in 1979, for whom the fertility rates are known up to the age of 40. As for the average age at childbirth, it fell overall from 1901 (29.4 years) to 1977 (26.5 years). Since then, it has been rising constantly, reaching 30.7 years in 2019.

Figure V – **Total fertility rate (on the left) and completed fertility (on the right) according to different assumptions**



Reading Note: In France, the total fertility rate is 1.87 children per woman in 2019. Completed fertility for women born in 1920 is an average of 2.5 children.
Sources and coverage: Insee, population estimates and civil status registry statistics from 1920 to 2019; Insee, population projections from 2013 to 2070. Metropolitan France for years up to 1993, France excluding Mayotte from 1994 to 2013, France from 2014 onwards.

The majority of experts have approved a ceiling on the average age at childbirth at 32 years, a total fertility rate stable at 1.95 and completed fertility of close to 2 children per woman. A total fertility rate stable at 1.95 with a ceiling of 32 years for the average age at childbirth results in completed fertility of 2.06 for the generations born between 1990 and 2005 and 1.95 for the generations born in 2020 and beyond (Figure V). In practice, the fertility rates are continued at each age according to the trend observed between 2009 and 2013. The ceiling for the average age at childbirth (32 years) is reached in 2040. A slight correction coefficient is applied for each year until 2040 in order to set the total fertility rate at 1.95, the target value approved by the experts. From 2040, the age-specific fertility rates are kept constant until 2070.

The low and high assumptions differ from the central assumption only in respect of fertility intensity and not its timing. While there was a broad consensus to have low and high assumptions for symmetrical total fertility rate compared to the central assumption, there was some debate on the setting of the bounds of the variants. We used + or - 0.15 children compared to the central assumption, which makes it possible to use the generation replacement threshold (2.1) as the high value, with the low assumption being a total fertility rate of 1.80 (Figure V). A fertility assumption in line with the EU average, with a total fertility rate of 1.6, was also constructed. In practice, within these variants, the total fertility rate reaches its target value in 2020 and stabilises after that date. In 2019, the total fertility rate is 1.87 children per woman in France, which fits between the low (1.80) and central (1.95) assumptions.

1.3. Migration

As in previous projection exercises, the migration assumptions relate to net migration by sex and age. This is measured indirectly by the difference between the population change between two successive censuses and the natural balance (births – deaths), using data taken from the civil status registry:

$$Net\ Migration_n = (Pop_{n+1} - Pop_n) - (Natural\ Balance_n)$$

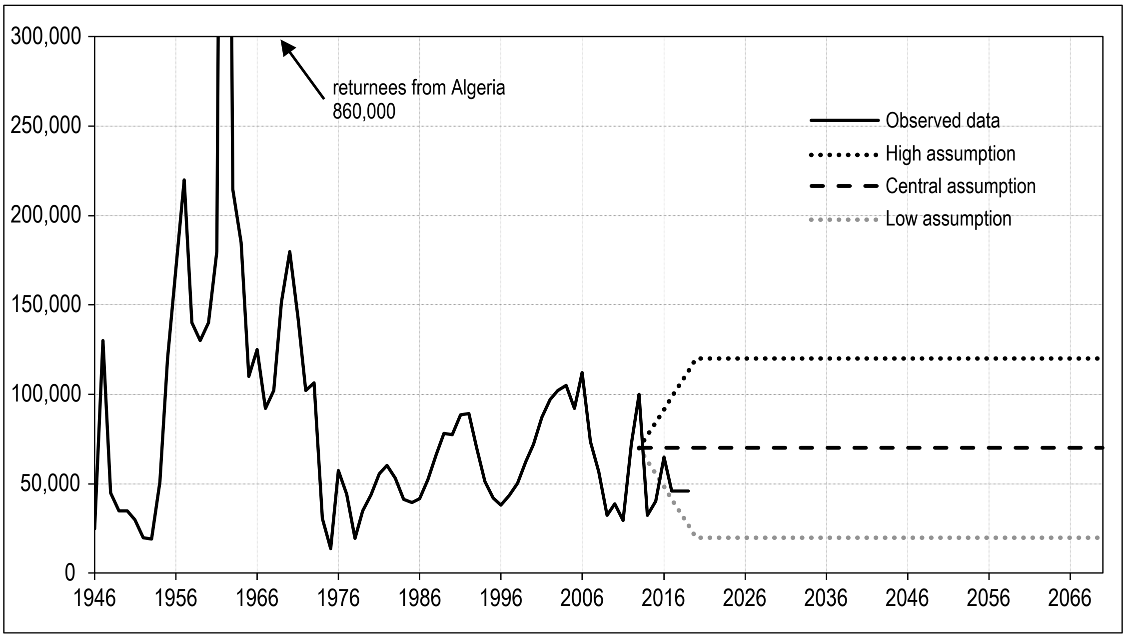
Until the 1980s, the central assumption of the projections reflected the “stated or assumed choices of the planner or migration policy”: an assumption based on the objectives of the economic development plans in the 1960s and 1970s, then the assumption of zero net migration in the 1979 and 1986 projections, in line with the policy of closing the borders to immigration from 1973 (Blanchet & Le Gallo, 2014). Subsequent projections are based more on past trends, allowing for results closer to the observed data. The central migration assumption of this projection exercise uses net migration of 70,000 people per year. This level is fairly close to the average calculated over different past periods (Figure VI). The structure by sex and age is assumed to be stable and corresponds to the average observed over the 2006-2012 period. However, some experts have highlighted the value of modifying this method by focusing on the flows of immigrants and emigrants by sex and age and no longer on net migration. Indeed, net migration is the result of movements of populations that are very diverse in terms of their motivations and migration history, their age and their profile at the time of migration. The people who make up the flow of immigrants are foreigners on arrival, who have various statuses (students, refugees, spouses of French nationals, etc.), in addition to French citizens returning or coming to live in France, whether they were born abroad or left to live abroad. On the emigration side, again, the motivations and ages are diverse. Unfortunately, it has not been possible to fully take into account this recommendation. Indeed, the flow of immigrants by sex and age is known in annual census surveys, thanks to a question on previous place of residence. In contrast, there are no comprehensive statistics that allow for the direct recording of flows of emigrants (Brutel, 2015). Emigration can only be estimated based on the difference between immigration and net migration:

$$Exits_n = Entries_n - Net\ migration_n$$

Emigration therefore combines uncertainties associated with the estimation of immigration and those associated with net migration, which makes it difficult to breakdown by sex and age. The assumptions therefore relate to net migration by sex and age, the figures for which are more robust than emigration by sex and age.

Compared to the central assumption, the low and high assumptions differ by 50,000 people per year upwards or downwards (Figure VI). Net migration would therefore be between 20,000 and 120,000 people per year. It varies greatly from year to year, but has remained within this range between 1979 and 2016 (Figure VI). In 2016, the last year for which figures are known, net migration is 65,000 people (Papon & Beaumel, 2020), which is close to the central assumption (70,000).

Figure VI – **Net migration according to different assumptions**



Reading Note: In France, net migration is 46,000 people in 2019.
Sources and coverage: Insee, population estimates and civil status registry statistics from 1946 to 2019; Insee, population projections from 2013 to 2070. Metropolitan France for years up to 1993, France excluding Mayotte from 1994 to 2013, France from 2014 onwards.

2. Analysis of the Projections: Robustness and Fragility

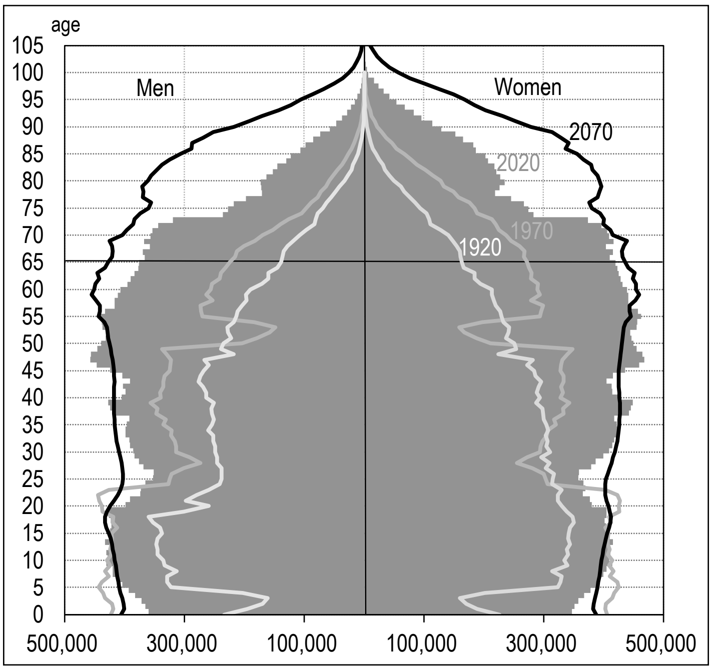
If recent demographic trends were to continue, France would have 76.4 million inhabitants in 2070, which is 9.4 million more than in 2020 (Table 1). Most of this increase would come from the elderly, defined here as those aged 65 or over (+8.2 million). This ageing of the population is not a new phenomenon. At the beginning of the 20th century, the age pyramid was aptly named: its base was wide and its top was pointed. It has gradually changed and now looks more like an "age cylinder" (Pison, 2009 and Figure VII). Indeed, the number of elderly has almost doubled every 50 years: 3.5 million elderly in 1920, 6.5 million in 1970 and 13.8 million in 2020. It could reach 21.9 million in 2070, according to the central scenario. However, the rate of growth up to 2070 would be lower than in the past: the number of elderly would increase "only" by a factor of 1.6 between 2020 and 2070, whereas it increased by a factor of 2.1 between 1970 and 2020 and 1.8 between 1920 and 1970. This

Table 1 – **Population and proportions by age in 1920, 1970, 2020 and 2070 (central scenario)**

	Metropolitan France			France		Metropolitan France		France
	1920	1970	2020	2020	2070	Evolution 1970/1920	Evolution 2020/1970	Evolution 2070/2020
Population (in thousands)								
Aged 0-19	11,999	16,748	15,390	16,085	16,262	40%	-8%	1%
Aged 20-64	22,841	27,306	36,055	37,228	38,243	20%	32%	3%
Aged 65 or over	3,543	6,474	13,453	13,751	21,944	83%	108%	60%
Total	38,383	50,528	64,898	67,064	76,448	32%	28%	14%
Proportion (as a %)								
Aged 0-19	31	33	24	24	21	6%	-28%	-11%
Aged 20-64	60	54	56	56	50	-9%	3%	-10%
Aged 65 or over	9	13	21	21	29	39%	62%	40%
Total	100	100	100	100	100			
Youth indicator (Aged 20-64/65 or over)	*6.4*	*4.2*	*2.7*	*2.7*	*1.7*	*-35%*	*-36%*	*-36%*

Reading Note: In 2070, France is expected to have 21,944,000 inhabitants aged 65 or over, according to the central scenario.
Sources and coverage: Insee, population estimates and civil status registry statistics in 1920, 1970 and 2020; Insee, central population projection scenario in 2070. Metropolitan France in 1920, 1970 and 2020, France in 2020 and 2070.

Figure VII – **Age pyramid for France in 1920, 1970, 2020 and 2070 (central scenario)**



Reading Note: In 2020, France has 419,000 women aged 65.
Sources and coverage: Insee, population estimates and civil status registry statistics in 1920, 1970 and 2020; Insee, central population projection scenario in 2070. Metropolitan France in 1920 and 1970, France in 2020 and 2070.

increase since 1920 is mainly the consequence of the increase in life expectancy. Each individual is more likely to become an elderly person than an individual of the generation born fifty years earlier. For example, 45% of men born in 1905 reached the age of 65 (in 1970), 76% of men born in 1954 reached this age in 2019 and virtually all (95%) men born in 2005 could live to become elderly in 2070.

In order to study ageing, it is necessary to look not only at the elderly, but also at younger people: indeed, the population ages if the number of young people increases less quickly than the number of elderly. A traditional indicator is the number of people aged between 20 and 64, which largely corresponds to the working ages, compared to the number of elderly, which mainly covers retired people. This ratio has been declining since 1920, indicating that the number of people aged 20 to 64 is increasing less quickly than the number of elderly and, therefore, that the population is ageing: from 6.4 people aged 20 to 64 per elderly person in 1920, the ratio fell to 4.2 in 1970, then 2.7 in 2020 and could be 1.7 in 2070 (Table 1).

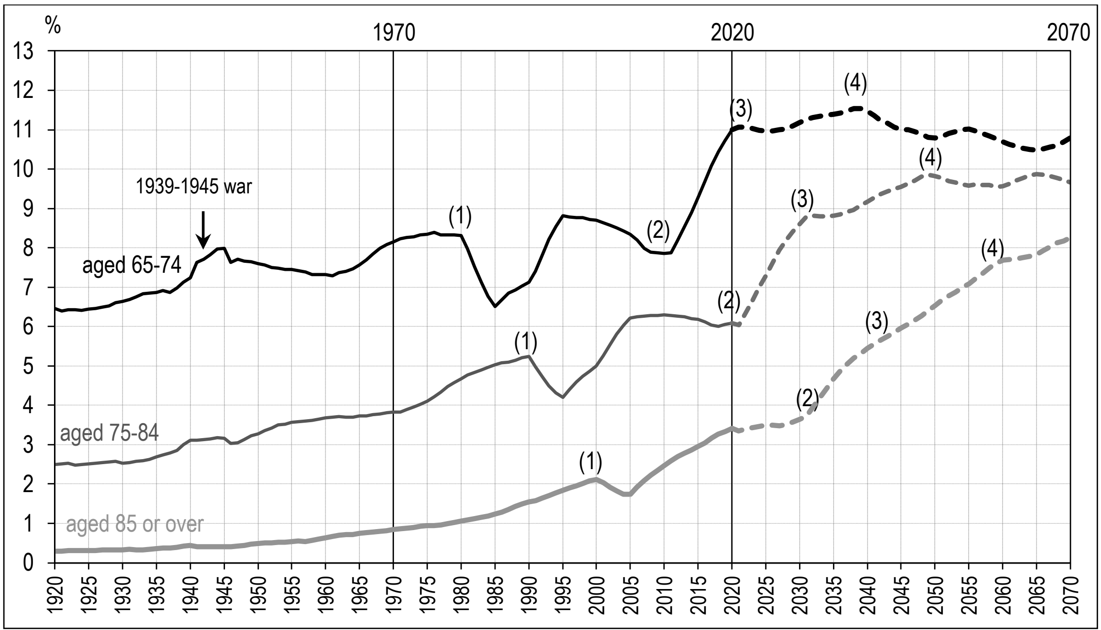
The rate of ageing, measured by the decline in the ratio of people aged 20 to 64 to those aged 65 or over, is expected to be similar over the next 50 years to that observed in the past (-36%, see Table 1). Some of the baby boom generations have already become elderly before 2020 (those aged between 65 and 73 on 1 January 2020). In contrast, the increase in the number of people aged 65 and over is expected to slow from 2040 onwards, by which time the last generation of the baby boomers will be over 65.

2.1. Between Now and 2070, Population Ageing Driven by the Eldest

The proportion of "young" elderly, aged 65 to 74, is expected to be virtually stable until 2070, close to 11% over the entire period (Figure VIII). It has increased since 2011, when the larger baby boom generations, born between 1946 and 1974, began to reach 65 years of age. From 2021, the people aged 65 to 74 will all have been born after the baby boom and the proportion of them within the population is expected to change little.

Only the eldest, aged 75 or over, are expected to contribute to the ageing of the population, as the first baby boom generation has not yet reached this age in 2020. The increase in the proportion of people aged 75 to 84 within the population is therefore expected to accelerate from 2021, with the increase in the proportion of people aged 85 years or older accelerating from 2031. Once each age group includes only generations born after the start of the baby boom, the ageing is expected to continue due to the rise in life expectancy, but at a slower rate, up to 2050 for those aged 75 to 84 (at which point their proportion within the population is expected to

Figure VIII – **Proportion of elderly people by age group and year**



Reading Note: In 2020, France has 11% of its population aged 65 to 74.
Notes: (1) Start of the arrival of the baby bust generations born 1915-1919; (2) Start of the arrival of the baby boom generations; (3) End of the arrival of the baby boom generations; (4) Death of the baby boom generations.
Sources and coverage: Insee, population estimates and civil status registry statistics from 1920 to 2020; Insee, central population projection scenario from 2021 onwards. Metropolitan France for years up to 1990, France excluding Mayotte from 1991 to 2013, France from 2014 onwards.

reach 9.8%) and up to 2060 for those aged 85 or over (at which point their proportion within the population is expected to reach 7.7%). Next, the effect of the increase in life expectancy on ageing should slow down with the death of the last baby boom generations: the proportion of people aged 75 to 84 is expected to stabilise at the end of the period (at 9.7%) and that of those aged 85 or older is expected to continue to rise (up to 8.2%).

The increase in life expectancy for more than a century in France has been accompanied by a coming together of ages at death. Under the 1920 mortality conditions, ages at death are highly variable: for women, 10% of deaths occur before the age of 1 year, 80% between the ages of 1 and 84 and 10% after the age of 84 (Table 2). Therefore, the age range within which 80% of deaths occur is 83 years. Under the 1970 mortality conditions, that age range is just 34 years, with 10% of deaths occurring before the age of 57 and 10% after the age of 91. This coming together of the ages at death has been achieved in particular thanks to an especially marked drop in mortality between birth and the age of 35. This is still continuing today: year after year, on average, deaths are occurring later and later and at ever closer ages (Figure IX). According to the central projection scenario, this coming together will continue: under the 2070 mortality conditions for women, 80% of deaths are expected to take place between the ages of 83 and 102.

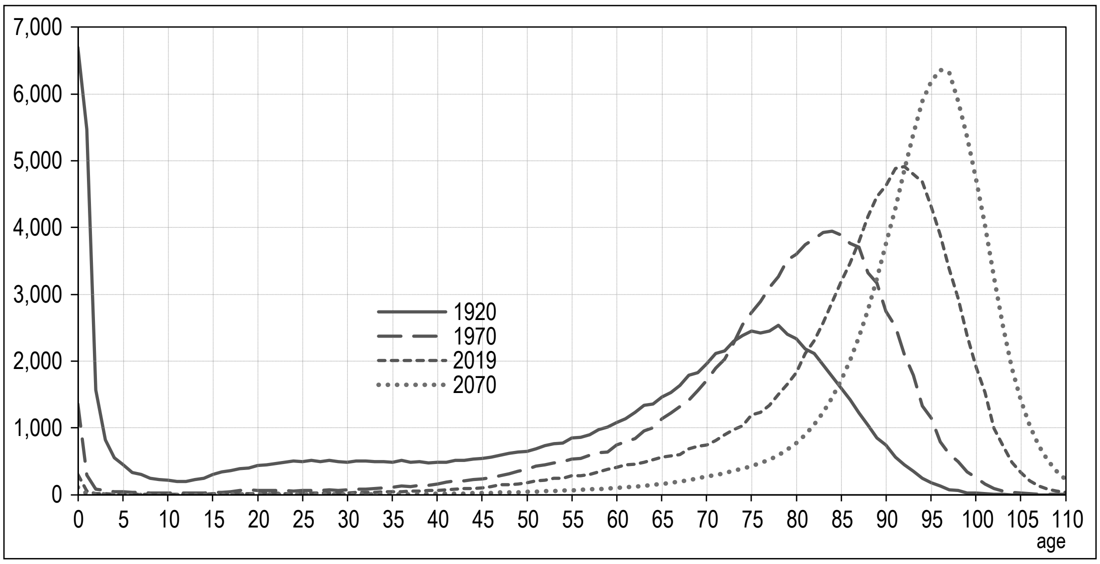
Table 2 – **Age before which 10%, 50% and 90% of men or women would have died under the mortality conditions of a given year**

(In years)

	Men				Women			
	10%	50%	90%	Interdecile range	10%	50%	90%	Interdecile range
1920	1	60	81	80	1	65	84	83
1970	47	72	87	40	57	80	91	34
2019	60	84	95	35	69	89	98	29
2070	78	92	100	22	83	95	102	19

Notes: These are deaths for a fictitious generation subject to the mortality conditions of a given year.
Reading Note: Under the 1920 mortality conditions, 10% of men would die before the age of 1 and 90% would die before the age of 81, giving an interdecile range of 80 years.
Sources and coverage: Insee, population estimates and civil status registry statistics from 1920 to 2019; Insee, central population projection scenario in 2070. Metropolitan France in 1920 and 1970, France in 2019 and 1970.

Figure IX – **Distribution of women's deaths under mortality conditions of a given year, per 100,000 deaths**



Reading Note: Under the 2019 female mortality conditions, 4,900 deaths would have occurred at the age of 92 (out of a total of 100,000 deaths).
Sources and coverage: Insee, population estimates and civil status registry statistics from 1920 to 2019; Insee, central population projection scenario from 2070 onwards. Metropolitan France in 1920 and 1970, France in 2019 and 2070.

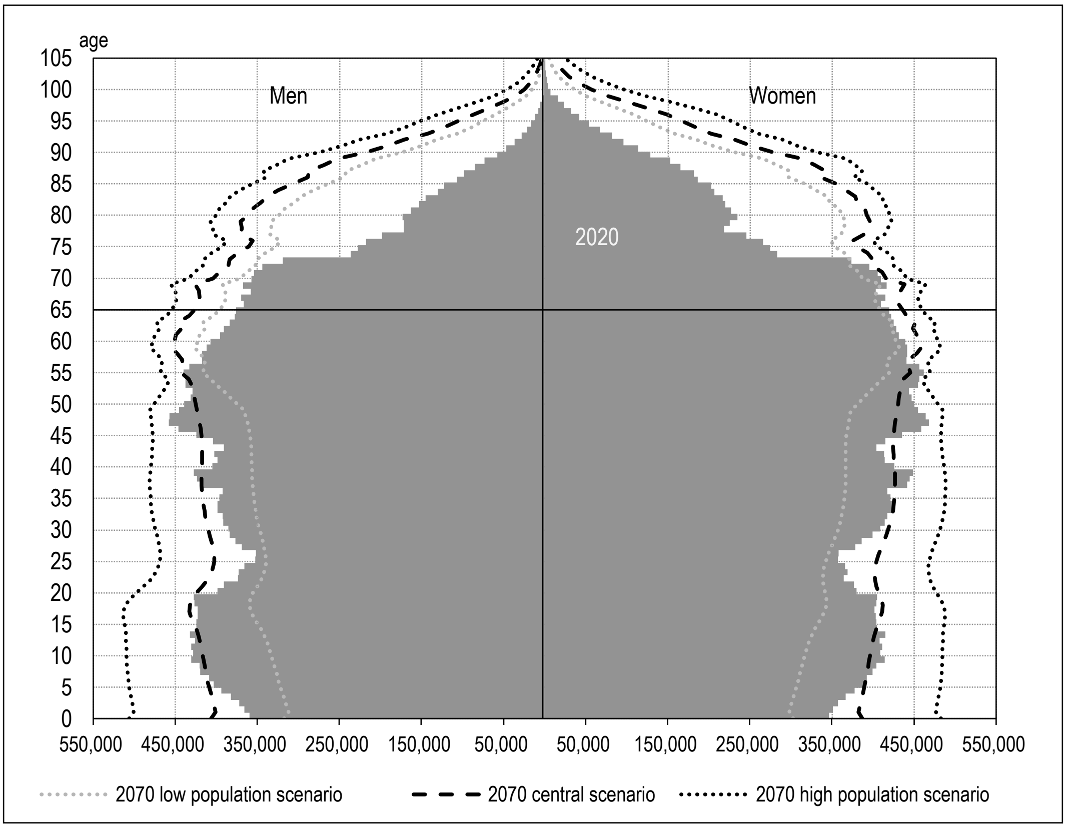
For men, the ages at death have also been moving closer together since 1920 and this trend is also expected to continue. For example, the age range in which 80% of deaths occur has been reduced from 40 years under the 1970 mortality conditions to only 35 years under those of 2019. Furthermore, in 2019, the spread of ages at death is greater for men than for women, but this gap is expected to narrow by 2070.

2.2. Uncertainty over the Evolution of Population Figures

While ageing in the coming years seems inevitable, the size of the population is uncertain. This is especially true for people under the age of 55 in 2070, virtually all of whom have not yet been born (Figure X), and neither have the mothers of the babies of 2070 – only their grandmothers have been. The projection for the number of people under the age of 55 is based on the number of women of childbearing age, their emigration from and immigration to French territory, and the evolution of fertility rates. However, unlike mortality, which generally shows a downward trend, there is no real medium-term trend concerning fertility, at least in countries such as France, which completed their demographic transition several decades ago (Vallin, 2002). The future evolution of the total fertility rate is therefore difficult to estimate. According to Eurostat data, fertility has generally declined in recent years in countries that had high fertility rates, and in some cases the decline has been very fast. For example, Finland, one of the most fertile countries in Europe with a total fertility rate of 1.87 in 2010, is now below the European average with a total fertility rate of 1.41 in 2018 (OSF, 2019). In France, the total fertility rate has also fallen recently, but less sharply: it fell from 2.0 in 2010 to 1.86 in 2019 for France excluding Mayotte (Beaumel & Papon, 2020).

In 2070, depending on whether all the assumptions are combined downwards or upwards, the number of people aged under 55 is expected to be between 38.3 million and 53.3 million, which is between -16% and +17% compared to the central scenario (Table 3). Births are expected to number between 643,000 and 1,013,000, which is -21% and +24% compared to the central scenario. If France were to have a lower fertility level in the future, close to the European average, this would lead to 35.9 million people aged under 55, which is -21% compared to the central scenario. There is less uncertainty over the total number of people aged 55 or older than over the number of people who have not yet reached that age. Those aged over 55 in 2070 have already been born, as they are the people aged under 60 at present who will survive until that date and will stay or settle in France. The number of people aged 55 or over would be between 27.8 million and 34.4 million, which is between -10% and +11% compared to the central scenario, depending on whether all assumptions are combined downwards or upwards. Only the scenario in which life expectancy remains at its 2014 level would lead to a more significant change, -20% compared to the central scenario.

Figure X – **Age pyramid for France in 2020 and 2070**
Central scenario, low and high population scenarios



Reading Note: In 2020, France has 419,000 women aged 65.
Sources and coverage: Insee, population estimates and civil status registry statistics in 2020; Insee, population projections in 2070. France in 2020 and 2070.

Table 3 – **Population by age (in millions) for various scenarios in 2070 and difference (in %) from the central scenario**

		Low population	Central scenario	High population	Constant life expectancy	EU fertility	Immigration equal to emigration
Aged 0-54	Population	38.3	45.6	53.3	45.0	35.9	41.3
	Difference	-16%		17%	-1%	-21%	-9%
Aged 55 or over	Population	27.8	30.9	34.4	24.8	30.9	28.5
	Difference	-10%		11%	-20%	0%	-8%
Aged 55-64	Population	8.4	8.9	9.4	8.4	8.9	8.0
	Difference	-6%		6%	-6%	0%	-11%
Aged 65-74	Population	7.6	8.2	8.8	7.3	8.2	7.7
	Difference	-7%		7%	-11%	0%	-7%
Aged 75-84	Population	6.7	7.4	8.1	5.8	7.4	6.8
	Difference	-9%		9%	-22%	0%	-8%
Aged 85-94	Population	4.2	5.1	6.1	2.9	5.1	4.8
	Difference	-17%		21%	-43%	0%	-5%
Aged 95 or over	Population	0.8	1.2	2.0	0.4	1.2	1.2
	Difference	-32%		60%	-66%	0%	1%
Total	Population	66.1	76.4	87.6	69.8	66.8	69.8
	Difference	-14%		15%	-9%	-13%	-9%

Notes: In the EU fertility scenario, the total fertility rate of 1.6 children per woman from 2020 onwards.
Reading Note: According to the high population scenario, France would have 53.3 million inhabitants aged 54 or under in 2070.
Sources and coverage: Insee, population projections in 2070. France.

As for the total population residing in France, how it will evolve is uncertain. According to the low population scenario, the population would increase until around 2040, before decreasing and ending up just slightly higher in 2070 than in 2020 (Figure XI). In contrast, according to the high population scenario, the population would maintain a strong growth rate and reach 87.6 million in 2070, which is 20.6 million higher than in 2020.

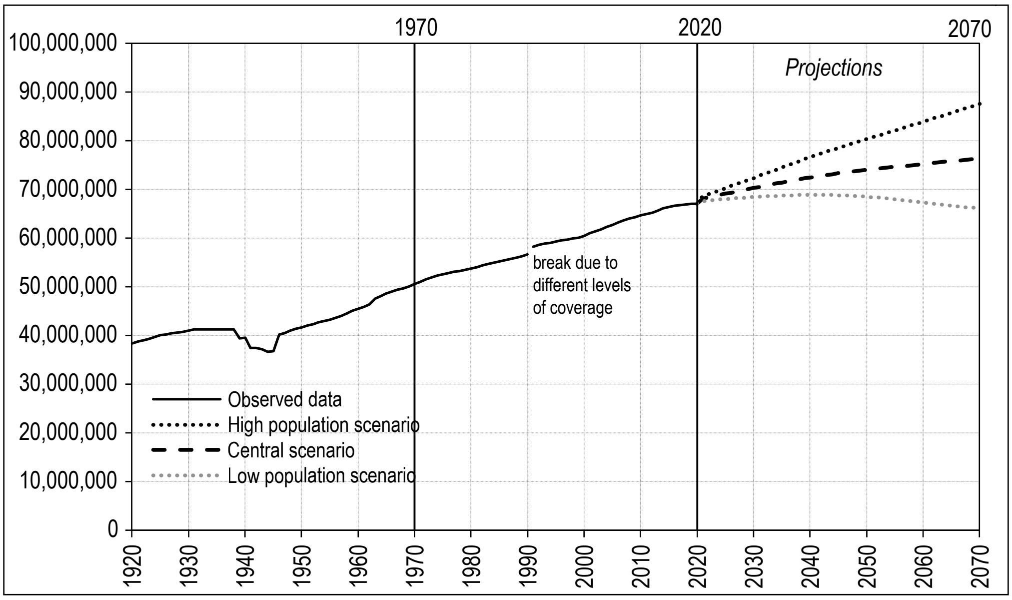
The central population projection scenario assumes that past trends will continue. Life expectancy at birth, for men, would then increase from 80 years in 2019 to 90 years in 2070, while for women it would increase from 86 to 93 years. To what extent does ageing depend on assumptions regarding life expectancy? To answer this question, we can analyse what would happen if life expectancy were not to increase. We assume that it remains at its 2014 level until 2070. However, in such a case, the population would age between 2020 and 2040: the difference compared to the central scenario and the scenario with constant life expectancy is relatively small (Figure XII). The proportion of elderly people would then increase from 20.5% to 24.5%, which is an increase fairly close to that of the central scenario (from 20.5% to 26.1%). Similarly, the ratio between the number of people aged 20 to 64 and those aged 65 or over would fall from 2.7 to 2.2 in 2040, compared to a decrease from 2.7 to 2.0 in the central scenario (Figure XIII). Thus, until 2040, ageing depends relatively little on the expected improvements in life expectancy. This is mainly a consequence of the past, i.e. the improvement of life expectancy that has already occurred and the continuation of the numerous baby boom generations living beyond the age of 65.

Beyond 2040, the constant life expectancy scenario does not call into question the increase in the number of elderly people aged 65 or over, but the assumptions used play a greater role. In 2070, the difference between the central scenario and the constant life expectancy scenario is more marked than in 2040 (Figure XII). Similarly, the evolution of the ratio between the number of people aged 20 to 64 and the number of elderly people is sensitive to the selection of assumptions: it would stabilise if life expectancy remained at its 2014 level, while it would decrease in the central scenario, albeit at a slower rate than in the past (Figure XIII).

2.3. The Ageing of the Population Depends on the Indicator Used

To study ageing, chronological age is often used, with a given fixed threshold, such as 65 years, for example. Another approach, using "prospective" age, i.e. the number of years left to live rather than the number of years already lived (Sanderson & Schervov, 2007), has been developed in particular in Belgium (Vandresse, 2020) and in Great Britain (Spijker & MacInnes, 2013).

Figure XI – **Observed and projected population under different scenarios**

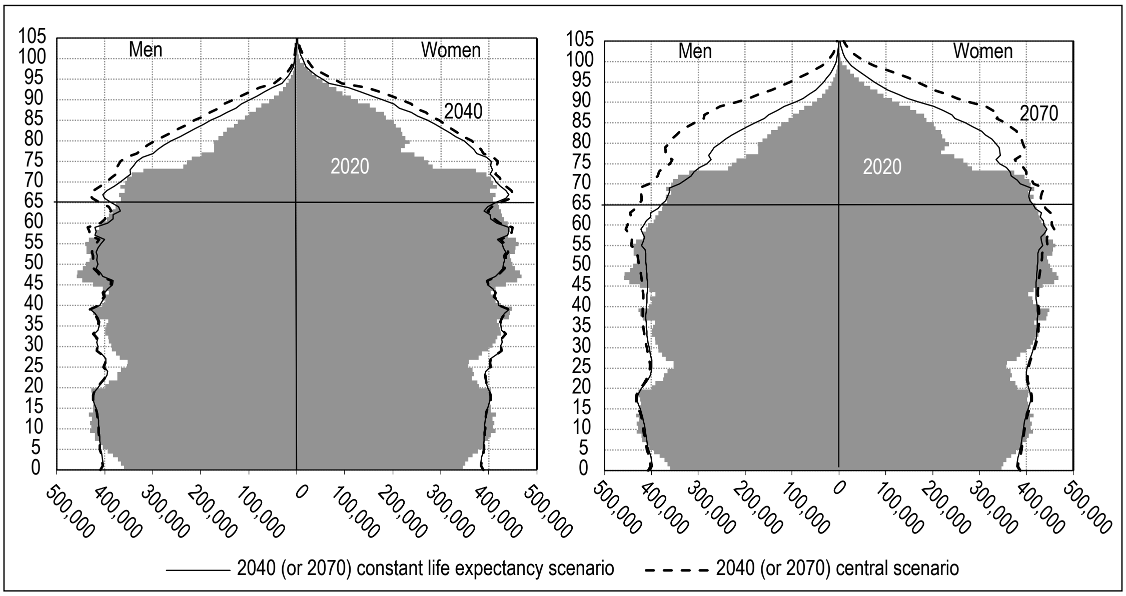


Reading Note: According to the central projection scenario, France would have 76.4 million inhabitants in 2070.
Sources and coverage: Insee, population estimates and civil status registry statistics from 1901 to 2020; Insee, population projections from 2021 onwards. Metropolitan France for years up to 1990, France excluding Mayotte from 1991 to 2013, France from 2014 onwards.

The previous analysis using chronological age, with a threshold at age 65, shows that the population of France has aged and that this phenomenon is expected to continue until 2070. What is the result when using prospective age? In this approach, the ageing indicator is calculated by dividing the number of people aged between 20 and the age at which life expectancy is 22 years by the number of people who are over that age and who, therefore, have a life expectancy of less than 22 years:[7]

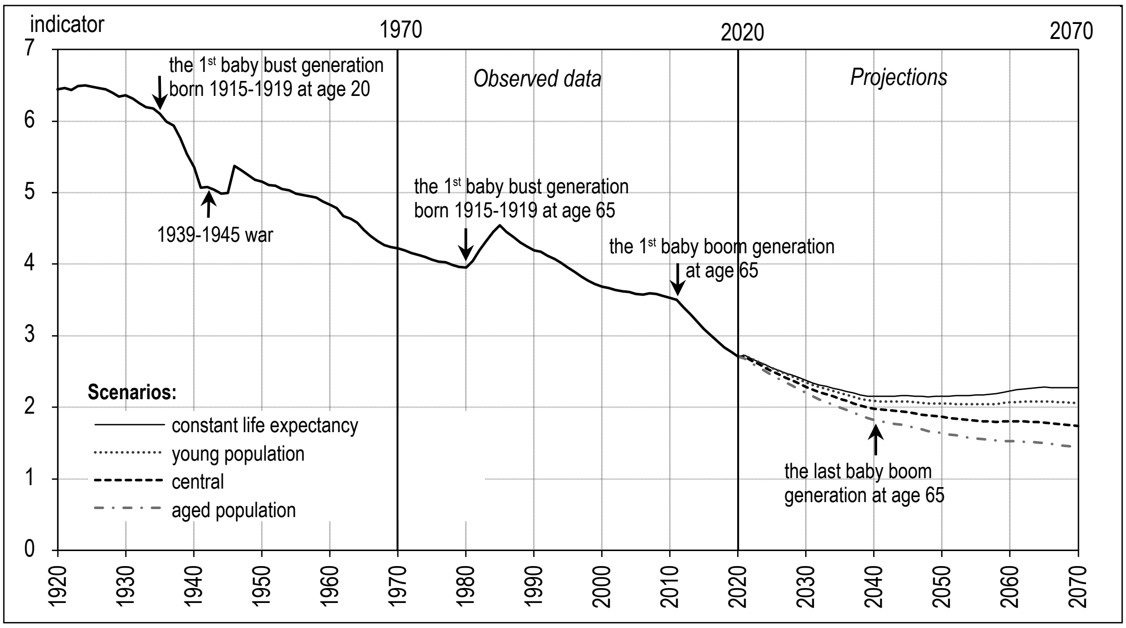
7. The threshold of 22 years has been chosen as that is the life expectancy at the age of 65 in France in 2019.

Figure XII – **Age pyramid in 2020 and 2040, and in 2020 and 2070**
Central scenario and constant life expectancy scenario



Reading Note: In 2020, France has 419,000 women aged 65.
Sources and coverage: Insee, population estimates and civil status registry statistics in 2020; Insee, population projections in 2040 and 2070. France.

Figure XIII – **Youth indicator[(a)]**



(a) Ratio of people aged 20 to 64 to elderly people aged 65 or over.
Notes: The young population scenario combines the assumptions of low life expectancy, high fertility and high migration, whereas the aged population scenario combines the opposite assumptions.
Reading Note: According to the central projection scenario, France would have 1.7 people aged 20 to 64 for each elderly person in 2070.
Sources and coverage: Insee, population estimates and civil status registry statistics from 1901 to 2020; Insee, population projections from 2021 onwards. Metropolitan France for years up to 1990, France excluding Mayotte from 1991 to 2013, France from 2014 onwards.

Using chronological age: $\frac{Pop\ 20\ to\ 64\ years\ old}{Pop\ 65\ years\ or\ over}$

Using prospective age: $\frac{Pop\ 20\ to\ \mathrm{x}\ years\ old}{Pop\ \mathrm{x}\ years\ or\ over}$

where *x* is the exact age (in years and months) at which life expectancy is 22 years for men or women. The age *x* therefore varies by year and sex.

Using this indicator, France has "become younger" since 1920: it had 1.8 people with a life expectancy of over 22 years per person with a lower life expectancy (Table 4). This ratio reached 2.1 in 1970 and 2.8 in 2020. This becoming younger can be explained by the strong increase in the number of people with a life expectancy of over 22 years, combined with a slight increase in the number of people with a lower life expectancy. By 2070, according to the prospective approach, France should neither become younger nor older: the ratio would be 2.7, which is almost the same level as in 2020.

Thus, the ageing of the population depends on the indicator used. Using the number of years lived, France will age and is expected to continue to age according to the central scenario. Using the number of years left to live, France has become younger and should neither age nor become younger between now and 2070. The selection of the most appropriate indicator depends on the purpose of the study and the assumptions used. For example, when studying the evolution of the number of people of dependent age, chronological age will be the more appropriate indicator if healthy life expectancy is assumed to be stable, whereas prospective age will correspond better to healthy life expectancy assumed to evolve at the same rate as life expectancy.

3. France and its EU Neighbours

The population projections published by Eurostat (Eurostat, 2019) make it possible to compare France's situation with that of its European neighbours. These projections are not a simple compilation of the national projections made by each country, but are a different exercise, with a common methodology for all countries of the 28-member EU, as well as for Iceland, Norway and Switzerland.[8] The advantages of selecting a common method rather than compiling national projections are: the absence of missing data for some countries that do not yet produce projections; easy access to documentation and results; and the elimination of bias associated with each country's varying degrees of optimism, which facilitates comparisons. The downside of this method is the inevitable discrepancy between Eurostat's projections and those carried out by the countries' national institutes. This discrepancy can lead to communication problems and questions regarding the data to be selected by users.

Like Insee, Eurostat uses the component method and a deterministic approach to establish the reference scenario. For each of the components, Eurostat uses as a basis a continuation of past trends and an assumption of convergence of demographic dynamics within Europe, which is based on the idea that socio-economic differences between EU countries are bound to narrow. As time progresses, the use of the continuation of past trends gives way to the use of the assumption of convergence. Convergence is partially achieved in 2100, the European projection horizon. The total fertility rate is thus projected to rise everywhere except in France, the country with the highest fertility in 2018, where it remains almost stable. The total fertility rate increases more in low-fertility countries, allowing for convergence. Life expectancy is projected to rise in all countries, with those with low life expectancy making more gains than others. Net migration increases in countries where it is negative and falls in countries where it is strongly positive, which also allows for convergence.

8. United Nations (UN) projections are also based on a common methodology, rather than a compilation of national projections. They use a probabilistic approach (Costemalle, 2020).

Table 4 – **Population (in thousands) in 1920, 1970 and 2020 and evolutions according to prospective age**

	Metropolitan France			France		Metropolitan France		France
	1920	1970	2020	2020	2070	1970/1920	2020/1970	2070/2020
Population aged 20 to *x* (a)	17,085	22,915	36,377	37,558	44,083	34%	59%	17%
Population aged *x* or older (b)	9,300	10,865	13,131	13,421	16,104	17%	21%	20%
Total	26,384	33,780	49,508	50,979	60,187	28%	47%	18%
Youth indicator (a/b)	1.8	2.1	2.8	2.8	2.7	15%	31%	-2%

Reading Note: In 2070, France has 44,083,000 inhabitants aged 20 to *x*, with *x* being the exact age at which life expectancy is 22 years for men or women.

Sources and coverage: Insee, population estimates and civil status registry statistics in 1920, 1970 and 2020; Insee, central population projection scenario in 2070. Metropolitan France in 1920, 1970 and 2020, France in 2020 and 2070; people aged 20 or over.

Together with the reference scenario, Eurostat provides a scenario with zero net migration for each year projected (with fertility and mortality assumptions identical to those of the reference scenario), so as to better understand the population evolution mechanisms linked to migration.

In their reference scenario, Insee and Eurostat make very similar assumptions about net migration, of around 70,000 people per year on average over the period 2019-2069. The total fertility rate projected by Eurostat for France (1.87 in 2070) is slightly lower than Insee's central assumption (1.95), but remains higher than the low assumption (1.80). It is in respect of mortality that the differences are most marked: according to the Eurostat reference scenario, life expectancy at birth would reach 86.6 years for men and 91.0 years for women in 2070, which is a level close to the low Insee hypothesis (87.1 year and 90.0 years). The difference is linked to the fact that Eurostat carried out its projections more recently than Insee and has thus been able to take greater account of the slowdown in the rise in life expectancy observed since 2014.

According to Eurostat, the 28-member EU would have 509.5 million inhabitants in 2070, only slightly less (-0.8%) than its 513 million inhabitants in 2019 (Table 5). Initially, the population would increase slightly until 2044 (+2.2%), then would decrease to its initial level at the end of the period.

Table 5 – **Population, net migration, total fertility rate and life expectancy at birth according to country of residence**

	Population in 2019 (millions)	Population in 2070 (millions)	Evolution 2019/2070 (%)	Net migration/population 2019-2069 (%)	Total fertility rate in 2070	Life expectancy	
						Men in 2070	Women in 2070
Luxembourg	0.6	1.0	68	0.8	1.62	86.6	90.7
Malta	0.5	0.7	47	0.8	1.61	86.8	90.7
Sweden	10.2	14.5	42	0.5	1.81	86.7	90.1
Cyprus	0.9	1.2	33	0.5	1.53	86.4	89.7
Iceland	0.4	0.5	30	0.3	1.76	86.9	90.2
Ireland	4.9	6.1	25	0.2	1.79	86.7	90.3
UK	66.6	82.1	23	0.3	1.81	86.3	89.9
Norway	5.3	6.5	22	0.4	1.69	86.8	90.3
Switzerland	8.5	10.4	22	0.4	1.64	87.2	90.8
Denmark	5.8	6.6	14	0.3	1.79	86.1	89.8
Belgium	11.5	12.9	13	0.3	1.73	86.2	90.2
Austria	8.9	9.9	12	0.4	1.68	86.2	90.1
France	67.0	72.0	7	0.1	1.87	86.6	91.0
Spain	46.9	48.4	3	0.4	1.52	86.9	91.1
Netherlands	17.3	17.4	1	0.2	1.70	86.5	89.8
UE28	513.5	509.5	-0.8	0.2	N/A	N/A	N/A
Germany	83.0	80.6	-3	0.3	1.71	86.0	89.9
Czech Rep.	10.6	10.2	-4	0.2	1.77	84.8	89.1
Finland	5.5	5.3	-5	0.2	1.62	86.0	90.4
Slovenia	2.1	1.9	-9	0.2	1.74	85.7	90.1
Estonia	1.3	1.2	-13	0.1	1.76	84.2	89.6
Hungary	9.8	8.5	-13	0.2	1.74	83.6	88.4
Slovakia	5.5	4.6	-16	0.1	1.65	84.1	88.9
Poland	38.0	31.7	-17	0.0	1.67	84.3	89.4
Italy	60.4	50.2	-17	0.3	1.53	86.8	90.6
Greece	10.7	8.5	-21	0.1	1.56	86.3	90.1
Portugal	10.3	8.0	-22	0.1	1.56	85.8	90.3
Romania	19.4	15.1	-22	0.0	1.79	83.5	88.4
Latvia	1.9	1.4	-25	0.0	1.79	82.6	88.5
Croatia	4.1	2.9	-28	0.0	1.59	84.2	88.6
Bulgaria	7.0	4.8	-31	0.0	1.69	83.0	87.7
Lithuania	2.8	1.9	-32	-0.1	1.72	82.9	88.6

Reading Note: According to the Eurostat reference scenario, the 28-member EU would have 509.5 million inhabitants in 2070. In France, net migration compared to the population would be an average of 0.1% per year between 2019 and 2069.
Sources and coverage: Eurostat, demo_pop in 2019 and europop2018 in 2070. 28-Member EU and Iceland, Norway and Switzerland. France includes Mayotte and Saint-Martin.

Why would the population of the EU be virtually the same in 2070 as in 2019? Eurostat projects an increase in the total fertility rate between 2019 and 2070 for all countries except France, where the total fertility rate would remain virtually stable. Nevertheless, it remains below the generational replacement threshold (2.1 children per woman) for all countries and over the entire period. Fertility therefore has a decreasing influence on the evolution of the total population. In contrast, life expectancy would increase between 2019 and 2070 for all countries, which has an increasing influence on the evolution of the population. Do these two effects offset each other? To answer this question, Eurostat has developed a scenario with zero net migration, i.e. with a number of emigrants equal to the number of immigrants. According to this scenario, the EU would have 419.9 million inhabitants in 2070, which is a decrease of 18% compared to 2019 (Table 6). The increase in life expectancy would therefore not offset the fact that total fertility rate is below the generational replacement threshold. In contrast, in the reference scenario, Eurostat projects positive average net migration over the 2019-2069 period for almost all countries except Lithuania, Latvia and Romania. It would therefore be this migration that would partly explain the stability of the EU population. The migration would combine with increased life expectancy to compensate for low fertility. The virtual stability of the EU population masks disparities between countries. Some countries could see their populations grow, sometimes sharply, mainly those located in Northern or Western Europe, while others could see their population decrease, generally those located in the East (Table 5).

The EU population is expected to age by 2070: the number of elderly, driven by the increase in life expectancy, would increase sharply (+45%), while the number of younger people would fall, -8% for those under 20 and -14% for those aged 20 to 64. Therefore, the ratio between the number of people aged 20 to 64 and those aged 65 or over would fall: from 3.0 in 2019 to 1.8 in 2070. As in France, the ageing of the European population is not a new phenomenon. In the 27-member EU,[9] the ratio has thus fallen from 4.2 in 1990 to 3.0 in 2019. All countries in the 28-member EU, as well as Iceland, Switzerland and Norway would be affected by population ageing as a result of improved life expectancy combined with low fertility. Eurostat also projects an ageing of the population for France, with the ratio falling from 2.8 to 1.9. However, ageing is slightly more marked in the Eurostat projections (-30%) than

9. Croatia joined the EU in 2003.

Table 6 – **Population (in millions) by age and youth indicator in 2019 and 2070**

	2019	2070	Evolution 2070/2019
28-Member EU (Eurostat – reference scenario)	513.5	509.5	-1%
Aged 0-19	106.6	97.7	-8%
Aged 20-64	304.1	263.0	-14%
Aged 65 or over	102.8	148.8	45%
Youth indicator (Aged 20-64/65 or over)	3.0	1.8	-40%
28-Member EU (Eurostat – zero net migration scenario)	513.5	419.9	-18%
Aged 0-19	106.6	75.1	-30%
Aged 20-64	304.1	207.3	-32%
Aged 65 or over	102.8	137.4	34%
Youth indicator (Aged 20-64/65 or over)	3.0	1.5	-49%
France (Eurostat – reference scenario)	67.0	72.0	8%
Aged 0-19	16.2	15.3	-5%
Aged 20-64	37.3	37.4	0%
Aged 65 or over	13.5	19.3	43%
Youth indicator (Aged 20-64/65 or over)	2.8	1.9	-30%
France (Insee – central scenario)	67.0	76.4	14%
Aged 0-19	16.2	16.3	1%
Aged 20-64	37.3	38.2	2%
Aged 65 or over	13.5	21.9	63%
Youth indicator (Aged 20-64/65 or over)	2.8	1.7	-37%

Reading Note: According to the Eurostat reference scenario, the 28-member EU would have 509.5 million inhabitants in 2070.
Sources: Eurostat, demo_pop in 2019 and europop2018 in 2070 for the 28-member EU; Insee, population estimates and civil registry statistics in 2019 and population projections in 2070 for France.

in the Insee projections (-37%), mainly due to a lower life expectancy assumption (Table 6).

* *
*

The population projections make it possible to describe the long-term future of the population, under certain assumptions. Even though the central scenario of the population projections has no chance of happening exactly as established, it still provides a lot of information. The objective of a projection is to present the most likely assumptions within a range of possibilities. Among all of the scenarios provided, the central scenario is often preferred. The projections highlight this scenario, which continues past trends, and present alternative scenarios that would occur if the rate of evolution of the components were to speed up or slow down. The role of demographers is, in particular, to indicate which results differ greatly between scenarios and which vary only slightly. This role is also to highlight those scenarios that depend mostly on our past and little on our future. Certain demographic phenomena, such as the continued ageing of the population, are already included in the current age pyramid. By comparing different assumptions, the projections make it possible to understand the mechanisms that explain the future evolution of the population.

The benefits of a projection are therefore varied, despite the uncertainties inherent in the exercise, which can result in discrepancies between projections and observed evolutions. Various studies have compared the results of past projections with the actual data for France (Blanchet & Le Gallo, 2014) and for some European countries (Majerus, 2015). In France, for example, the population grew at a faster rate than projected in all exercises between 1986 and 2010. In contrast, the continued ageing of the population had already been anticipated. The ratio of people aged 20 to 59 to those aged 60 or over was projected to be close to its current level as early as 1986. These studies therefore teach us to be cautious and show the need to take into account the sensitivity of the results to different assumptions. ☐

BIBLIOGRAPHY

Beaumel, C. & Papon, S. (2020). Bilan démographique 2019. La fécondité se stabilise en France. *Insee Première* N° 1789. https://www.insee.fr/fr/information/4286182

Blanchet, D. & Le Gallo, F. (2014). Retour vers le futur : trente ans de projections démographiques. Insee Références, *Trente ans de vie économique et sociale*, pp. 33–42.
https://www.insee.fr/fr/statistiques/1374368?sommaire=1374377

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/2400057

Blanpain, N. & Buisson, G. (2016b). Projections de population 2013-2070 pour la France. *Insee Résultats* N° 187-Société. https://www.insee.fr/fr/statistiques/2496793

Brutel, C. (2015). L'analyse des flux migratoires entre la France et l'étranger entre 2006 et 2013. Un accroissement des mobilités. *Insee Analyses* N° 22. https://www.insee.fr/fr/statistiques/1521331

Costemalle, V. (2020). Projections probabilistes bayésiennes de la population pour la France. *Economie et statistique/Economics and Statistics*, ce numéro.

Desrivierre, D. (2017). D'ici 2050, la population augmenterait dans toutes les régions de métropole. *Insee Première* N° 1952. https://www.insee.fr/fr/statistiques/2867738

Équipes de surveillance de la grippe (2018). Surveillance de la grippe en France, saison 2017-2018. *Bulletin épidémiologique hebdomadaire*, 34, 664–674.
https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-et-infections-respiratoires/grippe/documents/article/surveillance-de-la-grippe-en-france-saison-2017-2018

Eurostat (2019). Summary methodology of the 2018-based population projections (EUROPOP2018).
https://ec.europa.eu/eurostat/cache/metadata/en/proj_esms.htm#stat_pres1573482062759

Hubert, M., Bunle, H. & Boverat, F. (1937). *La population de la France, son évolution et ses perspectives*. Paris: Hachette.

Keilman, N. (2020). Evaluating Probabilistic Population Forecasts. *Economie et statistique/Economics and Statistics*, ce numéro.

Koubi, M. (2017). Projections à l'horizon 2070. Une hausse moins soutenue du nombre d'actifs. *Insee Première* N° 1646. https://www.insee.fr/fr/statistiques/2844302

Majérus, P. (2015). Étude comparative des analyses ex post des projections démographiques luxembourgeoises, belges, françaises, britanniques et néerlandaises du vingtième siècle. STATEC, *Économie et statistiques* N°82/2015.
http://www.statistiques.public.lu/fr/publications/series/economie-statistiques/2015/82-2015/index.html

Meslé, F. (2006). Progrès récents de l'espérance de vie en France : les hommes comblent une partie de leur retard. *Population* 61(4). 437–462.
https://www.ined.fr/fr/publications/editions/conjoncture-demographique/progres-recents-de-l-esperance-de-vie-en-france-les-hommes-comblent-une-partie-de-leur-retard/

Official Statistics of Finland (2019). Births [e-publication]. Helsinki: Statistics Finland [referred: 23.7.2019]. http://www.stat.fi/til/synt/2018/synt_2018_2019-04-26_tie_001_en.html

Papon, S. (2019). La situation démographique en 2017. *Insee Résultats*.
https://www.insee.fr/fr/statistiques/4136000

Papon, S. & Beaumel C. (2020). Bilan démographique 2019. *Insee Première* N° 1789.
https://www.insee.fr/fr/statistiques/4281618#consulter

Pison, G. (2007). La population de la France en 2007. Ined, *Population et Sociétés* N° 443.
https://www.ined.fr/fr/publications/editions/population-et-societes/la-population-de-la-france-en-2007/

Pison, G. (2009). Le vieillissement démographique sera plus rapide au Sud qu'au Nord. Ined, *Population et Sociétés* N° 429.
https://www.ined.fr/fr/publications/editions/population-et-societes/le-vieillissement-demographique-sera-plus-rapide-au-sud-qu-au-nord/

Pison, G. (2017). France 2016 : la natalité est en baisse. Ined, *Population et Sociétés* N° 542.
https://www.ined.fr/fr/publications/editions/population-et-societes/france-2016-natalite-en-baisse/

Pison, G. (2019). Pourquoi l'espérance de vie augmente-t-elle moins vite en France ? Ined, *Population et Sociétés* N° 564.
https://www.ined.fr/fr/actualites/presse/pourquoi-lesperance-de-vie-augmente-t-elle-moins-vite-en-france/

Roussel, R. (2017). Personnes âgées dépendantes : les dépenses de prise en charge pourraient doubler en part de PIB d'ici à 2060. Drees, *Études et résultats* N° 1032.
https://drees.solidarites-sante.gouv.fr/etudes-et-statistiques/publications/etudes-et-resultats/article/personnes-agees-dependantes-les-depenses-de-prise-en-charge-pourraient-doubler

Sanderson, C. & Schervov, S. (2007). A new perspective on population ageing, *Demographic Research*, 16, 27–58. https://www.demographic-research.org/volumes/vol16/2/default.htm

Spijker, J. & MacInnes, J. (2013). Population ageing; the timebomb that isn't. *British Medical Journal*, 347:f6598. https://doi.org/10.1136/bmj.f6598

Vallin, J. (2002). The End of the Demographic Transition: Relief or Concern? *Population and Development Review*, 28(1), 105–120. http://www.jstor.org/stable/3092759

Vandresse, M. (2020). Âge prospectif : une approche complémentaire du vieillissement et de ses implications. *Gérontologie et société* vol 42/n°162, 141–159.

Wattelar, C. (2004). Perspectives démographiques : historique de la méthode et méthodes actuelles. *In:* Caselli, G., Vallin, J. & Wunsch, G. (Eds), *Démographie : Analyse et synthèse*, Tome V, pp. 253–264. Paris: Éditions de l'Ined.

Wicksell, S. (1926). Sveriges framtida befolkning under olika fôrutsattingar. *Ekonomisk Tidskrift*, 28(1), 91–123.

APPENDIX

EVOLUTION OF THE POPULATION OF FRANCE FROM 1935 TO 1985 ACCORDING TO A PROJECTION BY A. SAUVY

The projection by A. Sauvy

LA DÉPOPULATION A CRAINDRE 217

cette diminution conservent en effet toute leur puissance ou gagnent même en intensité.

Or, voici ce que nous montrent les calculs de M. Sauvy, en ce qui concerne le mouvement et le chiffre de la population :

	Naissances	Décès	Excédent des décès	Population
1935...	638.000	658.000	20.000	41.426.000
1940...	523.000	611.000	88.000	41.249.000
1945...	475.000	601.000	126.000	40.702.000
1950...	452.000	595.000	143.000	40.048.000
1955...	407.000	585.000	178.000	39.270.000
1960...	349.000	578.000	229.000	38.283.000
1965...	280.000	572.000	292.000	37.006.000
1970...	229.000	568.000	339.000	35.447.000
1975...	190.000	566.000	376.000	33.685.000
1980...	155.000	565.000	410.000	31.734.000
1985...	127.000	556.000	429.000	29.645.000

Perspectives en cas de fécondité et de mortalité décroissant au rythme actuel.

Au premier abord la diminution prévue pour le nombre des naissances peut sembler invraisemblable, mais il n'en est plus de même dès que l'on compare ce nombre, pour une année donnée, au nombre des mariages probables en l'année considérée, nombre qui correspond à peu près à celui des naissances enregistrées 25 ans plus tôt, divisé par 2,5 à 3. C'est plutôt le nombre des décès prévu pour 1985 qui pourrait sembler exagérément faible, car il correspond à une mortalité par âge réduite de 65 % pour les hommes et les femmes de moins de 50 ans.

En ce qui concerne la composition par âge de la popu-

Hubert, M., Bunle, H. & Boverat, F. (1937). *The Population of France, its Evolution and its Outlooks.* Paris: Hachette.

Work carried out based on the civil status registry and population estimates

Year	Live births	Deaths [a]	Natural balance	Population in the middle of the Year
1935	644,000	662,000	-18,000	41,550,000
1940	561,000	740,000	-179,000	40,690,000
1945	646,000	644,000	2,000	39,660,000
1950	862,000	534,000	328,000	41,829,000
1955	806,000	526,000	280,000	43,428,000
1960	820,000	521,000	299,000	45,684,000
1965	866,000	544,000	322,000	48,758,000
1970	850,000	542,000	308,000	50,772,000
1975	745,000	560,000	185,000	52,699,000
1980	800,000	547,000	253,000	53,880,000
1985	768,000	552,000	216,000	55,284,000

[a] The number of deaths for the period 1939-1945 do not include deaths (civilian or military) by acts of war, i.e. approximately 600,000 people: 250,000 military personnel (regular army, prisoners of war and security forces) and 350,000 civilians (deported, shot and victims of land operations and bombings).
Sources and coverage: Insee, Statistics from the Civil Status Registry and Population Estimates. Metropolitan France.

Dimensions in Global Projections: An Overview

Anne Goujon*

Abstract – The addition of dimensions beyond age and sex in multistate population projections has two major objectives: first, to increase the accuracy of the projected population by capturing the heterogeneity present in the population that could affect the overall system; secondly, and more importantly, to increase the level of information provided by the projections. This article reviews the main dimensions that have been projected in the past, emphasizing global projections of educational attainment, which have been used largely in modeling exercises outside of the demographic realm. Furthermore, we propose some other dimensions that could be projected in a multistate fashion, possibly for most countries.

JEL Classification: J11, J24, I21
Keywords: population projections, multistate projections, education

* *European Commission Joint Research Centre, Ispra, Italy (anne.goujon@ec.europa.eu) and Wittgenstein Centre for Demography and Global Human Capital (Univ. Vienna, IIASA, OeAW/VID), Vienna Institute of Demography, Austrian Academy of Sciences, Vienna, Austria.*
This research work has been undertaken while the author was at the Wittgenstein Centre for Demography and Global Human Capital, and was finalized at the JRC.

Received July 2019, accepted July 2020.

Citation: Goujon, A. (2020). Dimensions in Global Projections: An Overview. *Economie et Statistique / Economics and Statistics*, 520-521, 87–101. https://doi.org/10.24187/ecostat.2020.520d.2032

Population projections have existed for a very long time. Beyond mere extrapolations, already several examples of population projections were carried out at the end of the 17th century with essays from John Graunt in 1662 (Graunt, 1665) and William Petty in 1682 (Petty, 1984), both projecting the population of London based on innovative statistical methods for their times.[1] In 1699, Sébastien Le Prestre de Vauban projected the Canadian population to 1970 – accurately despite false assumptions (Vauban, 1842). The projections that followed were greatly improved in terms of methodology. The cohort component method[2] (Whelpton, 1928) was developed in the 1920s and is widely used today. Nevertheless, most projections were for a long time not global. They were mostly implemented at the national or sub-national levels by national statistical offices and scientists.

We can speculate about several reasons for that. First of all, the absence of data for the base-year (needing a census or a survey) and for the fertility, mortality and migration components of the projections for a large number of countries, although many countries have been carrying out censuses already since the end of the 19th century. The second reason that might have been limiting the spread of global population projections is the computing capacity and time constraint to carry out the projections. However, most likely, the main reason might have been the absence of "global thinking", which came about with the emergence of the demographic transition theory, formulated in full by Notestein in 1945 but already elaborated by others (see Kirk, 1996). This theory, by assuming a continued global generalization of trends across countries, opened the door for global projections that were first developed by Notestein himself (Notestein, 1945). He became the first director of the United Nations Population Division, which for many decades was the main provider of global population projections. Then other large organizations joined in the production of global population projections, such as the World Bank, the Census Bureau, the Population Reference Bureau and the International Institute for Applied Systems Analysis (IIASA) to name the most prominent ones (see Lutz & KC, 2010 for a summary and timeline of global population projections and O'Neill *et al.*, 2001).

Global population projections are particularly needed for the inclusion of population in assessment models, usually as an exogenous variable, which helps to quantify the impact of the number of humans on other parameters. A relevant example of global population projection use is the work of the Intergovernmental Panel on Climate Change (IPCC), where population enters models that consider the vulnerability of populations to climate change, or that quantify economic activity by sector. Moreover, population being at the center of the development challenges of the coming century, it will affect the progress toward the realization of many sustainable development goals in 2030 and beyond, and therefore requires quantification.

For decades, global population projections have only included the dimensions of age and sex at the country level, mostly because there was no demand for more dimensions. Two research episodes revolutionized this apparent setting. In the 1980s, Andrei Rogers and a team of researchers working at IIASA developed the methodology of regional population projections (Rogers & Land, 1982). The researchers were concerned about taking into account demographic disparities between regions into a single projection model. A few years later, Nathan Keyfitz (1985) formalized the possibility to introduce additional dimensions in the projections, opening the door to a broader application of the multistate methodology. In short, the rational for adding dimensions to the projections follows the same rational as adding age and sex as dimensions in the projections, recognizing that the composition of the population can influence the results of the projection since different people have different demographic behaviour in terms of fertility, mortality, and migration. In other words, by adding granularity, the results of the population projections become more insightful, and secondly the projection results could be more exact by accounting for compositional effects in the projected population.

The multistate projection methodology relies on an extension of the cohort component method of population projection using the Leslie Matrix, as described in Keyfitz (1977) or Wunsch & Termote (1978). In the multistate extension, each Leslie matrix scalars for fertility and mortality are replaced by a matrix in each age group, which includes transitions between states. The transitions are one of the specificities of multistate projections that allow 'movements'

1. *Further back in time, Aristotle (384-322 BC) had already understood some of the principles of population projections as shown from this quote: "One would have thought that it was even more necessary to limit population than property; and that the limit should be fixed by calculating the chances of mortality in the children, and of sterility in married persons." (Book II, 1263b.15).*

2. *In short, the cohort component method divides the population to be projected into sex and age cohorts/groups to which are applied, year after year, different mortality, fertility, and migration rates.*

between states within the projection period, e.g. from primary education level to lower secondary level when states relate to educational attainment, or from rural to a urban areas when states relate to place of residence.[3]

In the first section of the paper, we will summarize briefly what dimensions have been projected. This section relies mostly on the literature that has been compiling such work. In the second section, we review the prerequisites to use a particular dimension as developed by Lutz *et al.* in 1998, and argue that some of the criteria could be relaxed and updated. In the third section, we suggest a few dimensions that could be projected and that could satisfy the criteria developed in section 2. In the final section, we discuss some of the challenges that producers of multidimensional projections should be aware of, before concluding.

1. What Have We Projected?

While multistate or multidimensional population projection[4] models are quite well known and well used nowadays, they are rarely implemented at the global level, where unidimensional population projections are still mostly being implemented. In an article in the *Philosophical Transactions of the Royal Society*, Lutz & KC (2010) reviewed some of the dimensions that have been projected at a global level, such as place of residence (e.g. United Nations 2018 for the latest round of projections from the United Nations), household composition (e.g. Habitat, 1996; Ironmonger *et al*., 2000), educational attainment (e.g. Lutz *et al*., 2018), marital status (e.g. Kantorová, 2013), religious affiliation (e.g. Pew Research Center, 2015), labor force participation (ILO, 2017 and 2018) and health (e.g. Global Burden of Disease Collaborative Network, 2016).

However, most of the dimensions above mentioned have not been projected in a multistate fashion, meaning that they do not fully model the demographic and dimensional interactions, and rely on a methodology based on prevalence often derived from econometric models (e.g. for labor force participation) or trend extrapolation. This is the case for instance of the United Nations urbanization prospects (United Nations, 2018) that provide population by place of residence up to 2050 for all countries. There are many difficulties in projecting place of residence, primarily because the definition of urban and rural zones is country-specific and changes over time. This brings an additional difficulty for multistate projections that model the mobility between urban and rural areas within the projections. This is also the case for other indicators, such as those related to global projections of poverty (e.g. Manuel *et al*., 2018).

Projections that attach prevalence rates to existing cohort-component projections usually do so in view of the difficulties to model the dynamic of the system, as mentioned in the case of place of residence. Another reason is that some dimensions are not very stable over the lifetime, as individuals might be mobile between dimensions. This is the case for place of residence but also for health status. Those types of projections usually assume scenarios with stable and changing prevalence/incidence rates over time and across regions, also modeling sometimes the risk factors affecting the dimensions.

The dimensions enumerated above particularly fit the list of criteria developed by Lutz *et al.* (1998) to be used to include a particular dimension in a projection, beyond age and sex. They were of three sorts:

1) The dimension should be "interesting in its own right and therefore desirable as an explicit output parameter" (Lutz *et al*., 1998 p. 42), giving precious information to the projection user. For instance, the number of one-person households, based on several dynamics such as patterns of divorce and leaving parental home, is an appealing parameter.

2) The dimension should be a source of demographic heterogeneity. It means that the fertility, mortality, and migration patterns of individuals should vary along that particular dimension. It is for instance the case with place of residence where the fertility of urban women tends in most cases to be much lower than that of rural women. This is linked to women's wider access to many resources such as family planning, education and health services, that would have a deterrent effect on their fertility, while at the same time adding constraints in terms of space availability to raise large families. Particularly, in low-income countries, the changing pattern of the differences will influence the future fertility depending largely on the urbanization rate. Impact of place of residence can also be found on mortality and on international migration. Education has also been

3. For a summary of the multistate methodology, see for instance Rogers (1981) or the Technical Note 1 in Goujon & Wils (1996).

4. In this paper, we do not differentiate between multi-dimensional and strict multistate projections (and use both terms interchangeably) where transitions between dimensions or states are either expressed as respectively probabilities or rates. We understand that the choice of one or the other will impede on the results, but this is not the purpose of the paper. Instead, we consider all population projection models where the population is decomposed further than by age and sex along one or more dimensions.

shown to have a large impact on demographic determinants, most of the time a negative effect on fertility, mortality and a rather positive effect on migration. The demographic heterogeneity introduced by the dimension, when taken into account, will have an impact on the dynamic of the system. For instance, Goujon & McNay (2003) and KC *et al*. (2018) have shown in the case of India the large impact of the granularity of the data in terms of state or place of residence and education.

3) While the first two criteria refer to the rationale behind adding a dimension to population projections, the third one is more practical and relates to the feasibility in terms of data availability (population, fertility, mortality, migration for each dimension, and transitions between the dimensions) and tools. Multistate population projections softwares have existed for some time: LIPRO,[5] which was developed originally for household projections[6] by the Netherlands Interdisciplinary Demographic Institute (NIDI), can also be used for a wide range of calculations in multistate demography. Additionally an R-package (MSDEM) is available for subnational multistate population projections.[7]

2. Population Projections by Levels of Education

Population projections of educational attainment are a rare case of global multistate population projections. They have been primarily developed at IIASA, starting with a first case study in the Mauritius Island (Lutz, 1994), followed by several applications at the national and regional level (e.g. Wils, 1996; Yousif *et al*., 1996; Goujon, 1997). In 2001, Goujon & Lutz (2004) projected for the first time population and education globally, for the world divided in 13 world regions. In 2010 were produced the first projections for a large number of countries, i.e. 120 countries, and four levels of educational attainment, to 2050 (KC *et al*., 2010). The number of countries was further increased to 171 in 2015, together with an increase in the number of categories to six, and a longer projection period up to 2100 (Lutz *et al*., 2014). The latest update was published in 2018 (Lutz *et al*., 2018; WIC, 2018[8]). The dataset now contains some 185 countries that comprised 99% of the world population in 2015. In the two latter exercises, the scenarios are based on both modeling and expert assessment about the future of fertility, mortality, migration, and education.

The assumptions about the projection are derived in two main steps (Lutz *et al*., 2014). First, expert opinion and models are used to derive the assumptions for the projection parameters overall, not taking into consideration levels of educational attainment, i.e. country level total fertility rates and age specific fertility rates, gender specific life expectancies and age and sex specific survival ratios, in and out migration rates and age and sex specific migration schedules. Country-specific education differentials are then obtained in a second step. For fertility, fertility levels by education for the base-year were obtained from the literature and from census and survey data. Countries with no available data were assumed to have the average fertility differentials of all countries from the broader region to which they belong. Education differentials are assumed to converge over time to certain ratios of TFRs for the different education levels relative to post-secondary education.[9] These values are assumed to be reached by the time TFR reaches 1.8 children per woman. For countries where the maximum differential is below 1.42 in the base-year, the relative ratios are then kept constant at those lower levels. The convergence hypothesis follows the literature showing that, in high-income societies, differentials become smaller in absolute and relative terms. Jalovaara *et al*. (2018) found that among the highly educated societies of Denmark, Norway and Sweden, there are almost to no differentials in the ultimate fertility of women between education categories[10] (see also Beaujouan & Berghammer, 2019).

For mortality, gender-specific education differentials in life expectancy at age 15 are standardized following findings in the literature. The difference in life expectancy at age 15 between the 'No education' category and the 'Post-secondary educated' population is assumed to be of six years for men and four years for women. Between these extreme points, a two-year difference is assumed between men with a completed primary and a completed lower secondary, and a one-year difference for the remaining levels of educational attainment. For

5. https://www.nidi.knaw.nl/en/research/al/270101 [accessed on 15/7/2019]
6. ProFamy is another existing software for projections of households and living arrangements available at http://profamy.com.cn/en_about.asp [accessed on 29/10/2019].
7. https://r-forge.r-project.org/R/?group_id=2281 [accessed on 15/7/2019]
8. The detailed dataset is available at http://dataexplorer.wittgensteincentre.org/ [accessed on 15/7/2019]
9. The ratios are 1.42 for women with no education, incomplete primary and completed primary education (≤ ISCED 1), 1.35 for women with lower secondary education (ISCED 2), 1.14 for women with upper-secondary education (ISCED 3), and 1 for women with post-secondary education (ISCED 4+) which is the reference category.
10. Interestingly, while highly educated mothers in Nordic countries reach more often higher birth rates at parity 2 and 3 compared to less educated ones, their completed fertility is nonetheless often slightly lower than that of women with less education, due to the later start of their childbearing career (Andersson et al*., 2009).*

women, the differential between the lowest and the highest education category is four years of life expectancy and proportionally split between education levels following the male division. The differentials are kept constant throughout the projection period. Finally, for children up to age 15, differential mortality is introduced through mothers' education.[11] For migration, where the lack of data on the characteristics of migrants is notorious, it is assumed that the education composition of migration flows is equal to that in the origin country.

The system is dynamic through a set of educational transition rates between education categories that are derived from national time series for all countries. These transitions occur between the ages of 15 to 34 years, considering that few people advance to a higher level of education after the age of 35. Because the model does not link individuals with their ancestry, the education transition of children does not depend on the education levels of parents.[12] These limitations among others are discussed in the penultimate section.

The main characteristics of these projections is that as mentioned above, when education is factored in, they tend to result in lower population growth than projections by age and sex. It is the main difference with the United Nations projections that lead to a world of 10.9 billion in 2100 in the medium variant (United Nations, 2019), compared to 9.3 billion according the trend scenario including the education dimension (WIC, 2018). This latter scenario also shows that most of the increase will occur at the level of the population with an upper- and post-secondary education level. This would mechanically affect fertility, which is overall much lower for the most educated categories (Figure I). For instance in Ethiopia in 2014-2016, the total fertility rate of women with no education or a primary education is 5.0 children compared with 2.1 children for those women with a secondary or higher education (according to the Demographic and Health Surveys[13]). While the scenario assumes that the differential gets smaller, in absolute terms, over the projection period, the momentum linked to large differentials has substantial consequences on the total population trends.

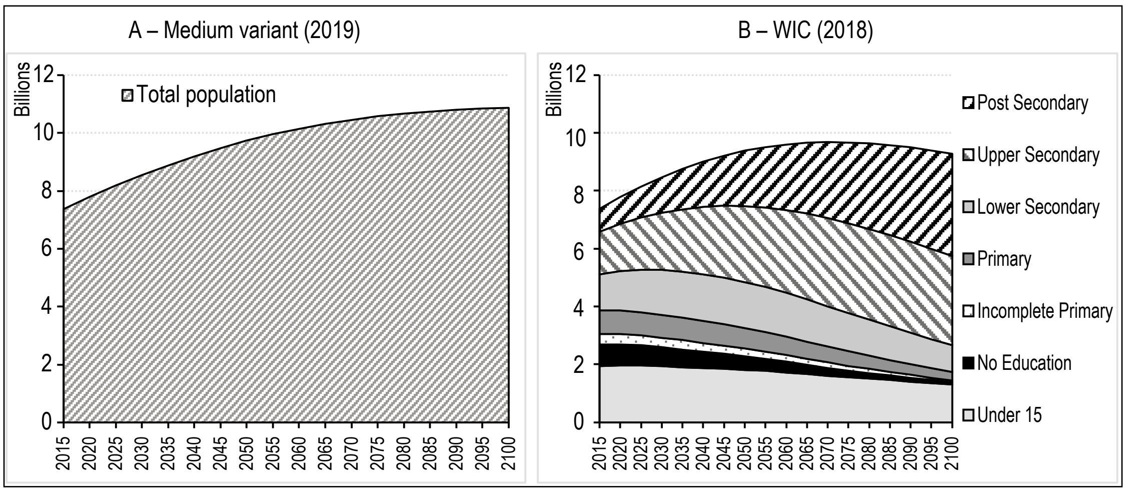
The projections of educational attainment have been applied by the modeling communities of the IPCC who have utilized the different scenarios to assess the relationships between socioeconomic development and climate change (KC & Lutz, 2014) and the role of education to reduce vulnerabilities and increase resilience (UNDP, 2014). They have also been employed to model the potential economic effect of future education paths in low-income countries (Basten & Crespo Cuaresma, 2014) and to model in general the link between education and economic growth (Lutz *et al.*, 2008). More recently, researchers have looked at the impact of education, particularly of women, in mitigating the labor market

11. The differentials in terms of relative ratio of mortality rates with respect to the completed upper-secondary category are 1.8, 1.7, 1.6, 1.4, 1.0 and 0.8, in ascending order of educational attainment – no education, incomplete primary education, completed primary education, completed lower secondary education, completed upper-secondary education, post-secondary education. These values are based on the averages of under-five mortality rates in the countries where Demographic and Health Surveys have been conducted.

12. In the framework of projections for 13 world regions, Goujon & Lutz (2004) calculated a scenario in which they incorporated a feedback from the level of education of mothers to the enrolment ratios for girls. This self-reinforcing mechanism has a positive impact on average levels of education but might also increase the dichotomy between the lowest – with little chance of moving up – and highest educated in the society.

13. Data from Demographic and Health Surveys *are available here: https://dhsprogram.com/ [accessed on 24/10/2019].*

Figure I – **Projections of the total world population by the United Nations**



Sources: United Nations (2019) and WIC (2018).

consequences of population ageing in the countries of the European Union (Marois *et al.*, 2019).

Although the strength of the dataset on educational attainment is that it relies on the collection and harmonization of existing data on education, one clear disadvantage is that it does not take into account the quality of education, which has been demonstrated to be very different across countries and also within countries for instance by Hanushek & Wößmann (2012). This requires further research in terms of data and modeling. In addition, the projections do not take into account constraints in terms of budget, infrastructure or work force associated with the development of education.

3. What Other Dimensions Have We Not Projected?

The above mentioned list of criteria that was developed to consider a particular dimension in multistate projections, and most notably to justify the inclusion of educational attainment (Lutz *et al.*, 1998) could be partly revised to increase the possibility of including more dimensions in the projections, especially when considering the impact outside the realm of demography. Indeed, that would be the case when a dimension is interesting and is a source of heterogeneity with an impact on the dynamics of the whole system, not necessarily related to demography heterogeneity as stated in the second criteria.

We develop below a list of potential dimensions that could be integrated into global population projections. We have restricted our list to dimensions that could be of interest at the global level – meaning not only for a specific population or for a country or region of the world – and that have not yet been projected at the global level, to the best knowledge of the author.

The rationale for selecting these particular dimensions is based on the following considerations:

- Their timeliness: these dimensions and related issues are present in the public debate and in the political agenda at the international level.
- Their generational (and gendered) features: as stipulated by the demographic metabolism theory (Lutz, 2013), societies change through generational replacement. The dimensions considered tend to be "sticky" along cohort lines, as for instance exemplified by studies and projections of the prevalence of a feeling of European identity in the European Union, and of changing attitudes towards homosexuality (Striessnig & Lutz, 2016a and 2016b).

There are some limitations to the proposed list of dimensions. First, this list does not pretend to be exhaustive and probably many other dimensions could be included. While these dimensions are interesting and their projections could inform about the potential consequences of some dynamics, they could also be seen as less robust as other dimensions such as education or place of residence. They are data intensive if one aims for global coverage and data availability has not been checked for all countries. Moreover, we do not develop in this paper the methodology that would be needed for the multistate population projections of these dimensions as such, assuming that they would be derived from the multistate methodology where most of the modeling needed would be about deriving the transition rates between the suggested states/dimensions.

3.1. Diet

The dimension what earth inhabitants will eat in the future is key to many factors affecting sustainable development. In this area, whether people have access to sufficient food is important.[14] However, beyond the adequacy of food supply, different nutrition behavior could be of importance and determine the ability of humans to live within the planetary boundaries (Rockström *et al.*, 2009). The share of the population that will adopt diets that are less rich in dairy and meat products such as vegan, vegetarian, or flexitarian diets has been shown to potentially have a significant impact on reducing the greenhouse gas emissions especially at the level of industrialized countries (Sandström *et al.*, 2018). Hence, dietary change could be an important tool "to limit global warming to less than 2°C, while providing a nutritious diet to a growing and changing world population" (Aleksandrowicz *et al.*, 2016, p. 1). While several studies have looked at what would be the impact of several dietary changes affecting climate change and the achievement of the Sustainable Development Goals, very few have considered how the propagation of dietary changes could happen in the population. This is particularly important because it is linked to individual characteristics such as age, gender, and possibly other background characteristics such as country of origin, place of residence, education and religion among others. It is also especially relevant for population projections because the changes will most likely follow a diffusion process across cohort lines, for instance

14. It is part of sustainable development goal 2: "End hunger, achieve food security and improved nutrition and promote sustainable agriculture".

from the rather young and more educated to the rest of the population.

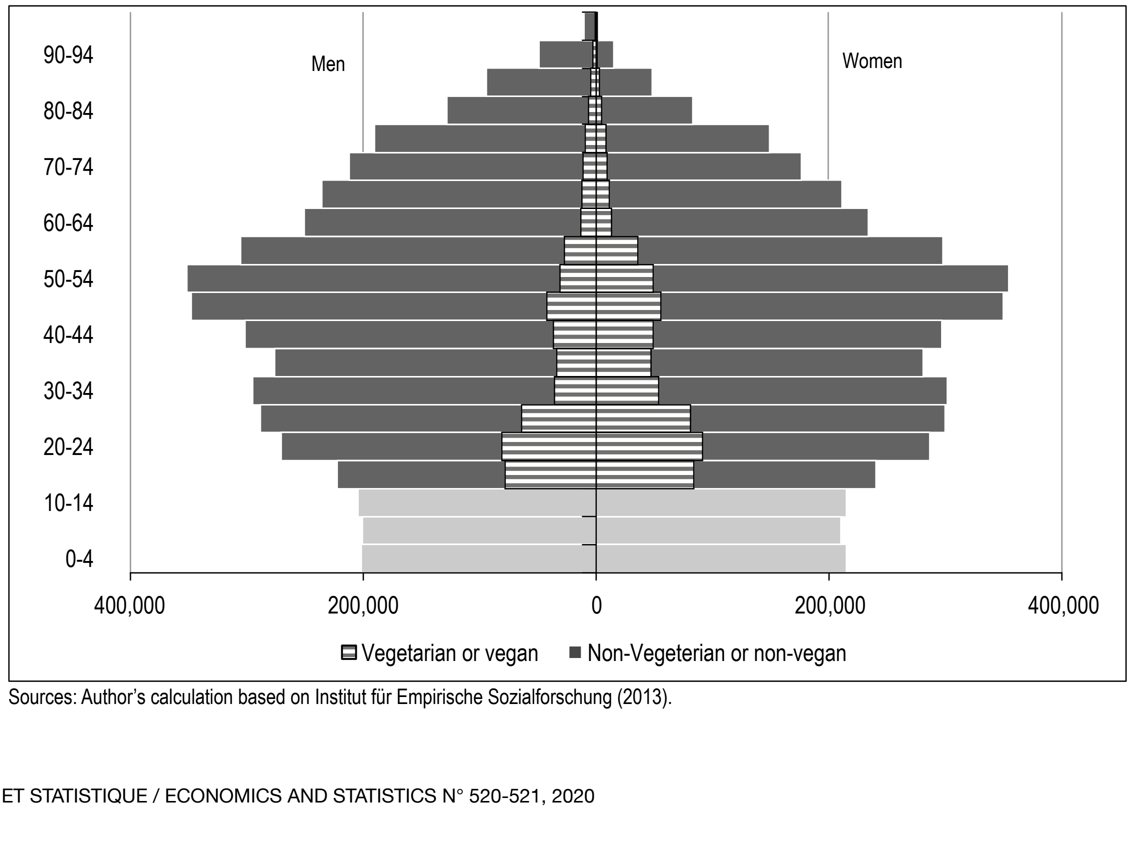
It would also be interesting to see what could be the impact on the demographic behavior. It has been shown for instance that vegan women suffer more often from amenorrhea when they do not supplement with vitamins such as B12 (Wokes *et al.*, 1955). It could also influence fertility for more environmentally concerned people, who often adopt a no- or less-meat diet and who are likely to want a small number of children (Arnocky *et al.*, 2011). The effect on fertility could be also mediated by other factors such as education, although the evidence is mixed on this topic (Allès *et al.*, 2017; Moreira & Padrão, 2004). The impact on mortality could also be substantial by reducing the prevalence of obesity and cardiovascular diseases in the population (Springmann *et al.*, 2018) and of some cancers associated with meat consumptions (Springmann *et al.*, 2016). All these phenomena would be interesting to consider in global population projections, also considering that several datasets on household expenditure surveys detail all the expenditures incurred by a large sample of individual households over a specified period (Leahy *et al.*, 2010), i.e. World Bank's Living Standard Measurement Studies (LSMS). The information is also available for some countries at the individual level, see for instance the estimates of the vegan population by age and sex shown in Figure II.

3.2. Language

While the implications of spoken languages might seem trivial in view of the potential challenges faced by the world population within the next century, it has some important implications at national or sub-national level. Size and concentration of language communities determine linguistic power, which in turn will influence the political power of those communities (Hung Ng & Deng, 2017). This can be seen in Canada (French and English), Belgium (Dutch, French and German) or in China (Mandarin, Cantonese and other languages such as Tibetan, Mongolian, etc.). Spoken languages will be influenced by the demographic vitality of the population speaking it. Internal and international migration would also play a major role in influencing this. While there is, evidently, no causal link between spoken language and demographic behavior, the variable of interest itself will be affected and can be projected using, implicitly or explicitly, other dimensions to determine the potential assumptions about the future demographic behavior of populations according to different languages. For instance, if Arab-speaking women in Israel were for a long time bearing more children than Hebrew-speaking ones in the rest of the country,[15] it is evidently not directly related to

15. This trend has been reversed in 2016 according to the Central Bureau of Statistics. While in 2002, the TFR of Arab women was 4.19 and that of Jewish women 2.64, in 2016, it is respectively 2.11 and 3.16. See https://old.cbs.gov.il/www/publications/lidot/lidot_all_1.pdf [accessed on 57/2019].

Figure II – **Estimated dietary preferences by age (from 15 years old) and sex in Austria in 2013**



Sources: Author's calculation based on Institut für Empirische Sozialforschung (2013).

the language but rather to the socio-economic conditions present in the region where these populations are concentrated on top of the political situation. While some researchers have been conducting language projections (e.g. Houle & Corbeil, 2017 and Sabourin & Belanger, 2015, for Canada; Ortman & Shin, 2011 for the United States of America), projections have not been carried out globally to see for instance the vitality of some languages (English or Chinese) as first or spoken languages for instance. Enumeration of population by languages is present in most censuses, whether they list native languages, home languages, and often the knowledge level of those languages, see for instance the distribution of the population in Finland by native language, at two points in time (Figure III). It is worth noting that the share of population whose native language is other than Finnish, Swedish and Sami has been noticeably increasing since 2000, particularly among the younger cohorts.

3.3. Political Allegiance and Ideology

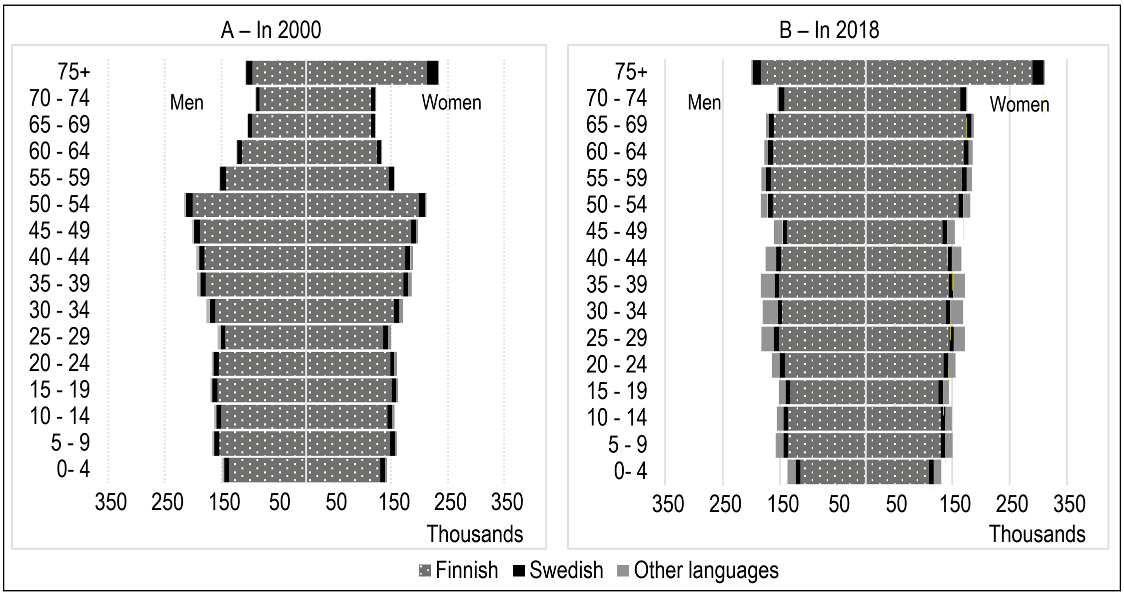
Few works have investigated the impact of differences in demographic behavior on socio-political variables, and even more so in a prospective manner, one exception being the work by Kaufmann *et al*. (2010) (Figure IV). However, in many societies, the tendency is for the electorate to cast increasingly their votes for populist parties (see Figure V). Research carried out for *The Guardian* estimates that the number of Europeans living under governments with at least one populist in cabinet has increased 13-fold between 1998 and 2018.[16] There are interesting demographic features about the voting behavior related, for instance, to age, gender (Harteveld *et al.*, 2015) and socio-economic characteristics (Rooduijn, 2018) especially education and place of residence, that could influence the future. Moreover, intergenerational transmission of ideology from parents to children (Jennings & Niemi, 1981; Abramowitz & Saunders, 1998; Jennings *et al*., 2009; Murray & Mulvaney, 2012) provides supplementary ground to study the dimension in a multistate manner in the sense that there is some stability in the system and less volatility than could be expected. Kaufmann *et al* (2010) present the rationale for projections of political ideology (distinguishing between liberals, moderates and conservatives) in the context of the United States of America, by stating that "if party allegiances are enduring and formed in early adulthood, much of the story of future American partisanship has already been written." (p. 12). However, it is not meant that the ideologies of the future depend solely on the demographic behavior of the population. "Pressures of the times" for young voters who first enter the electorate (Beck & Jennings, 1991, p. 742) and throughout their life time will also have an influence on determining the political ideology at the individual level.

3.4. Childlessness and Grand Childlessness

While many household projections exist and have looked at the composition of the households, few

16. *https://www.theguardian.com/world/ng-interactive/2018/nov/20/revealed-one-in-four-europeans-vote-populist [accessed on 17/7/2019]*

Figure III – **Population of Finland by age, sex and native language**

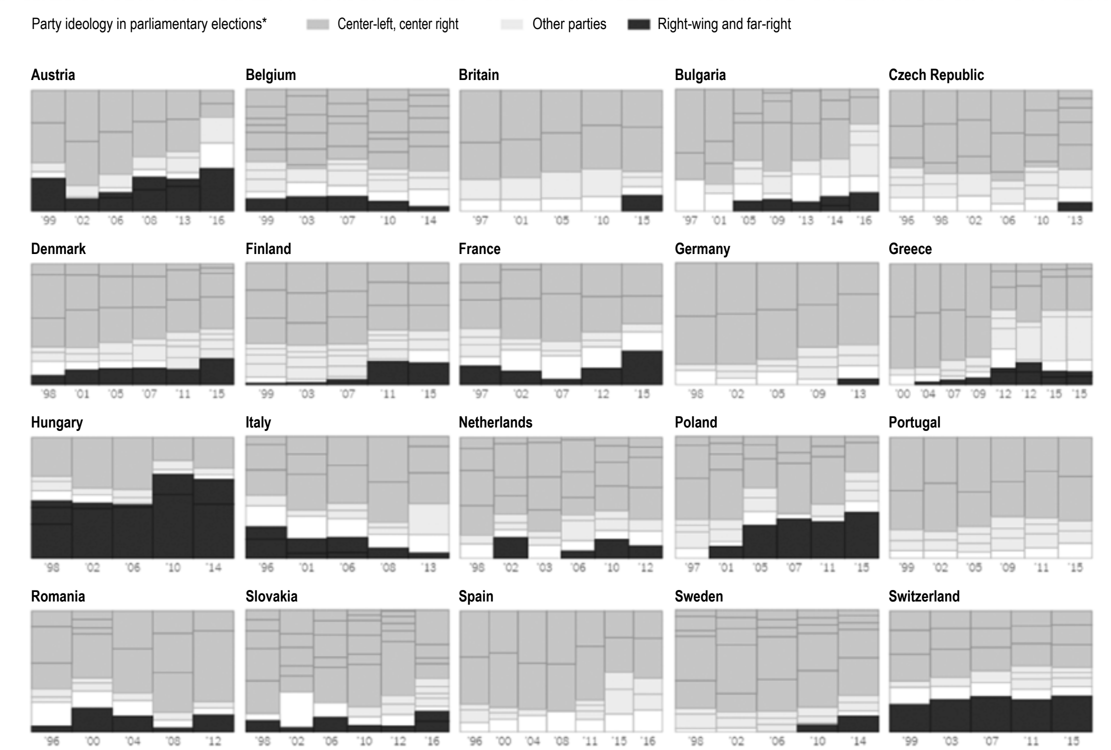


Sources: Author's calculation based on Population by age, sex and language, Statistics Finland (2018).

Figure IV – **Population pyramid of political allegiance in the United States, estimates (2003) and projections (2043)**

Notes: Children under age 21 inherit the political allegiance of their parents.
Sources: Kaufmann *et al.*, 2012 based on US General Social Surveys (2000-2006).

Figure V – **Party ideology in parliamentary elections, 1990s to 2010s, selected European countries**



*The 2016 presidential elections in Austria and Bulgaria and the 2015 French regional elections are included to add a more recent result for those countries.

Sources: New York Times (2016).
https://www.nytimes.com/interactive/2016/05/22/world/europe/europe-right-wing-austria-hungary.html

have looked at the changing repercussions of some recent trends across cohorts and generations. One interesting example of that is levels of childlessness that have been increasing in recent decades in Europe and in the Global North. It is particularly acute in Austria, Germany and Switzerland that are forerunners regarding this phenomenon, with more than 20% of women without children at the end of their reproductive career (Kreyenfeld & Konietzka, 2017). The occurrence of childlessness is spreading to other countries, first in Northern Europe, but also to Southern and Eastern Europe, and to East Asia.[17] Regardless of the causes explaining the absence of children into a man's or woman's life, this is likely to have consequences over their life course, particularly when they reach old age,

17. Childlessness is also on the increase in almost all world regions (United Nations, 2015).

missing that potential support. On the other hand, it is possible that elderly people do have children who do not have children themselves, therefore missing the experience of grand-parenting, which has implications in terms of not having descendants[18] (Margolis, 2016). This dimension could be studied along cohort lines as shown in Figure VI.

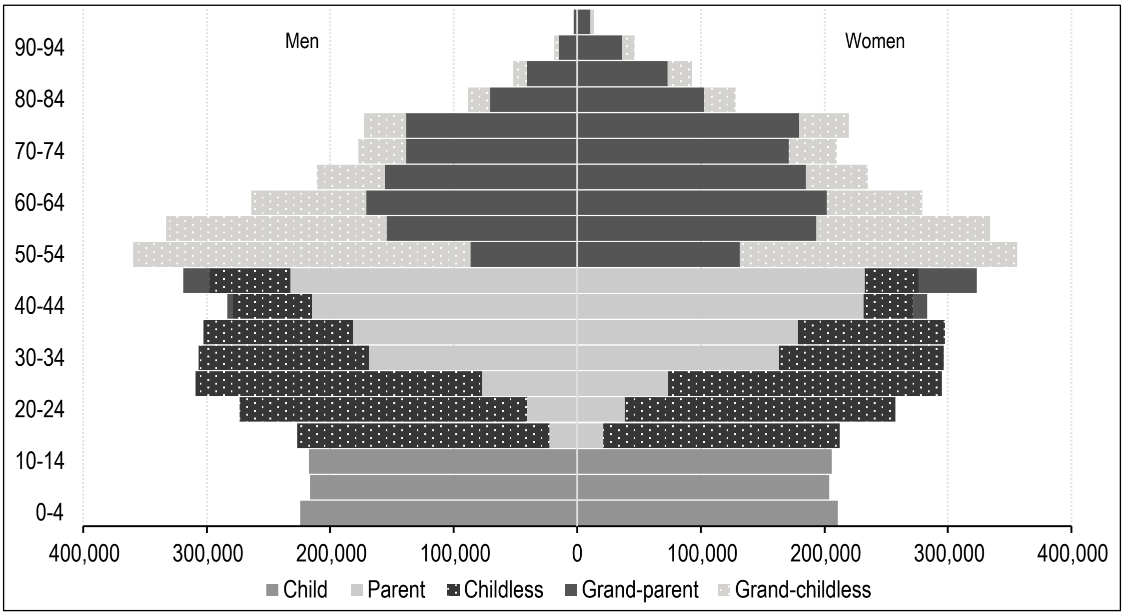
4. Potential Issues with Multistate Population Projections

Certain issues need to be considered when implementing multidimensional projections. First, a balance has to be reached between the number of variables that are necessary to improve the population projections results and the assumptions that will need to be made if more dimensions are considered. Indeed, with each additional dimension come a number of assumptions that have to be provided, related to the behavior of individuals in terms of fertility, mortality and migration. The availability of data for the base-year could become a limitation, especially when multiple dimensions are taken into consideration, e.g. education, place of residence and regions. While one can always revert to the assumption of no differentials in the absence of data, e.g., people in dimension 1 have the same fertility as those in dimension 2 and more, it would hamper the validity and relevance of the projections. Therefore, in developing multistate population projections, researchers need to use common sense to decide on the number of states. A possible compromise is to model the dimensions with existing data and a theoretical model and apply/model other population characteristics using prevalence rates without entering the projections as categories which, as we have shown, is being implemented in many forecasting exercises, e.g. projections of labor force participation based on multi-state projections of educational attainment (Loichinger & Marois, 2018).

Moreover, scenarios also model the relation between the chosen dimensions and the demographic determinants in the future, whose evolution can differ from what was observed in the past or in the present. For instance, as much as levels of educational attainment were and are a factor of heterogeneity explaining substantial parts of the changes in fertility across different countries, it is difficult to know what the role of education will be in the future, and what influence it will have on demographic behavior, assuming that most societies would be knowledge societies where information and knowledge would be the most important factor of production. Even when education still plays a major role, it will most likely not be the same education, as we

18. *Lineages without descendant disappear. As a result, people who have children but no grand children cannot become a numerically important category in the population. Especially when the variance of the number of children is low.*

Figure VI – **Hypothetical representation of childlessness and grand-childlessness by age and sex, in Austria in 2019**



Notes: The data presented on the graph are not real except for the age and sex structure of Austria in 2019. The distribution of the population between the different categories is simplified, assuming for instance that parenthood would be happening only between the ages of 15 to 49 years.
Sources: Author's concept.

understand it nowadays. In that sense, whether by adding granularity the results of the multistate population projections become more accurate depends highly on the ability of the model to predict changes in the link existing between the dimension and the demographic determinants. This caveat could be seen as deterrent to using multistate projections. However, we claim that it provides an opportunity to explore the sensitivity of the projections to different patterns of change of the relationship between the dimension and the demographic behavior of individuals along that dimension.

Not unrelated to the previous matter, the other challenge to be considered is that of causality that is underlying the projections at all time. While including the dimension in the projections influences the result because the dimension is a factor of heterogeneity, it does not necessarily mean that the dimension influences the demographic determinant in a causal way. A good example of that is the case of population projections of religious affiliation. In Europe, Muslim women's fertility is higher than that of Christian women for instance; however, this is not necessarily a direct effect of the religious affiliation but rather of the socio-economic background of these women within the different affiliations. When implementing a scenario, its interpretation has to be carefully formulated. Lutz & Skirbekk (2014) observe that "the assessment of causality in the social sciences is context-specific" (p. 18). They develop the idea that strong causality in intervention sciences, aimed at understanding "how the most important forces of change function in order to predict the future evolution of the system" (id., p. 18), is rather difficult to establish. On the other hand, social scientists should strive for functional causality – which differs from strong causality – that entails "strong empirically observed associations", supported by "plausible narratives about the mechanisms", and the elimination of "other obvious competing explanations" for the association observed between two factors (id., p. 19). They further show that, in that way, functional causality can be demonstrated from higher education to lower mortality and fertility "at least over the course of the demographic transition" (id., p. 28).

Another issue with multidimensional population projections models is the need to ensure consistency both internally (e.g. the problem of genders with projections of marital status) and externallly (e.g. when regional population projections should add up to national projections). Several algorithmic solutions exist to adjust each demographic component in order to minimize deviations (Keilman, 1985). Other research has looked at the issue of coherence particularly related to the modeling of future mortality patterns using the fact that differences between closely related populations are unlikely to increase in the long term. Therefore, projections of mortality (or of other determinants) for a sub-region or a sub-group could be improved by taking into account the patterns of a larger group (Li & Lee, 2005).

* *
*

The field of multidimensional projections is a thriving one. It is particularly active regarding education that has been projected in all kind of contexts and is used more and more in the global context as a proxy for development level, autonomy of women, and innovative and adaptive capacity. There are reasons to expect that the characteristics/dimensions of human beings will be interesting to project in a world that is keener on information about the future. It is also likely that studies about the future population will take more and more advantage of the availability of big data that could shed some lights on human behavior.

Like classical cohort-component projections, multistate population projections are more than forecasting tools as such, since they offer a tool to explore the future based on future assumptions looking at different scenarios based on "what if" narratives. In this sense, these scenarios look at the sensitivity of the projections to different assumptions. What those projections add to classical cohort-component projections is the influence and sensitivity of dimension that can be of importance for the projections itself. ☐

BIBLIOGRAPHY

Abramowitz, A. & Saunders, K. (1998). Ideological realignment in the US electorate. *Journal of Politics* 60(3), 634–652. https://doi.org/10.2307/2647642

Aleksandrowicz, L., Green, R., Joy, E. J. M., Smith, P. & Haines, A (2016). The Impacts of Dietary Change on Greenhouse Gas Emissions, Land Use, Water Use, and Health: A Systematic Review. *PLoS ONE* 11(11), e0165797. https://doi.org/10.1371/journal.pone.0165797

Allès, B., Baudry, J., Méjean, C., Touvier, M., Péneau, S., Hercberg, S. & Kesse-Guyot, E. (2017). Comparison of Sociodemographic and Nutritional Characteristics between Self-Reported Vegetarians, Vegans, and Meat-Eaters from the NutriNet-Santé Study. *Nutrients*, 9(9), 1023. https://doi.org/10.3390/nu9091023

Andersson, G., Knudsen, L. B., Neyer, G., Teschner, K., Rønsen, M., Lappegård, T., Skrede, K. & Vikat, A. (2009). Cohort fertility patterns in the Nordic countries. *Demographic research*, 20 (article 14): 313–352. https://dx.doi.org/10.4054/DemRes.2009.20.14

Arnocky, S., Dupuis, D. & Stroink, M. L. (2011). Environmental concern and fertility intentions among Canadian university students. *Population and Environment*, 34 (2), 279–292. https://doi.org/10.1007/s11111-011-0164-y

Basten, S. & Crespo Cuaresma, J. (2014). Modelling the macroeconomic impact of future trajectories of educational development in Least Developed Countries. *International Journal of Educational Development*, 36, 44–50. https://doi.org/10.1016/j.ijedudev.2013.12.003

Beaujouan, E. & Berghammer, C. (2019). The Gap between Lifetime Fertility Intentions and Completed Fertility in Europe and the United States: A Cohort Approach. *Population Research and Policy Review*, 38, 507–535. https://doi.org/10.1007/s11113-019-09516-3.

Beck, P. A. & Jennings, M. K. (1991). Family Traditions, Political Periods, and the Development of Partisan Orientations. *The Journal of Politics*, 53(3), 742–763. https://doi.org/10.2307/2131578

De Vauban, M. (1842). *Oisivetés de M. de Vauban, Tome IV, mémoire « Moyen de rétablir nos colonies de l'Amérique et de les accroître en peu de temps »*. J. Corréard: Paris.

Global Burden of Disease Collaborative Network (2016). *Global Burden of Disease Study 2016 Health-related Sustainable Development Goals (SDG) Indicators 1990-2030.* Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2017.

Goujon, A. (1997). Population and Education Prospects in the Western Mediterranean Region (Jordan, Lebanon, Syria, the West Bank and the Gaza Strip). *IIASA Interim Report* IR-97-046. Laxenburg, Austria: International Institute for Applied Systems Analysis. http://pure.iiasa.ac.at/5248

Goujon, A. & Lutz, W. (2004). Future Human Capital: Population projections by level of education. *In:* W. Lutz, W. C.Sanderson & S. Scherbov (Eds.), *The end of world population growth in the 21st Century*, pp. 121–157. London and Sterling (VA): Earthscan.

Goujon, A. & McNay, K. (2003). Projecting the educational composition of the population of India: Selected state-level perspectives. *Applied Population and Policy*, 1 (1), 25–35.

Goujon, A. & Wils, A. (1996). The Importance of Education in Future Population. Global Trends and Case Studies on Cape Verde, Sudan, and Tunisia. *IIASA Working Papers* WP-96-138. Laxenburg, Austria: International Institute for Applied Systems Analysis. http://pure.iiasa.ac.at/4889

Graunt, J. (1665). *National and Political Observations mentioned in a following index and made upon the Bills of Mortality (third edition)*. Royal Society, London.

Habitat (United Nations Centre for Human Settlements) (1996). *An Urbanizing World: Global Report on Human Settlements, 1996*. Oxford: Oxford University Press.

Hanushek, E. A. & Wößmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, 17, 267–321. https://doi.org/10.1007/s10887-012-9081-x

Harteveld, E., Van Der Brug, W., Dahlberg, S. & Kokkonen, A. (2015). The gender gap in populist radical-right voting: examining the demand side in Western and Eastern Europe. *Patterns of Prejudice*, 49:1-2, 103–134. https://doi.org/10.1080/0031322X.2015.1024399

Houle, R. & Corbeil, J.-P. (2017). Language Projections for Canada, 2011 to 2036. Ethnicity, Language and Immigration Thematic Series. Ottawa: Statistics Canada. https://www150.statcan.gc.ca/pub/89-657-x/89-657-x2017001-eng.pdf

ILO – International Labour Organization (2017). Labour force estimates and projections: 1990-2030 (2017 edition), Methodological description. Geneva: International Labour Organization

ILO – International Labour Organization (2018). Labour Force Estimates and Projections by age, July 2018 - Country data – Annual.
www.ilo.org/ilostat-files/WEB_bulk_download/modelled_estimates/LFEP_AGE_country.dta [accessed on 23/10/2019]

Ironmonger, D., Jennings, V. & Lloyd-Smith, B. (2000). Long Term Global Projections of Household Numbers and Size: Distributions for LINK Countries and Regions. Paper presented at the Project LINK meeting, Oslo, Norway on 3–6 October 2000.

Jalovaara, M., Neyer, G., Andersson, G., Dahlberg, J., Dommermuth, L., Fallesen, P. & Lappegård, T. (2018). Education, Gender, and Cohort Fertility in the Nordic Countries. *European Journal of Population*, 35(3), 563–586. https://doi.org/10.1007/s10680-018-9492-2

Jennings, M. K. & Niemi, R. G. (1981). *Generations and Politics: A Panel Study of Young Adults and their Parents*. Princeton, NJ: Princeton University Press.

Jennings, M. K., Stoker, L. & Bowers, J. (2009). Politics across Generations: Family Transmission Reexamined. *Journal of Politics*, 71 (3), 782–799. https://doi.org/10.1017/s0022381609090719

Kantorová, K. (2013). National, Regional and Global Estimates and Projections of the Number of Women Aged 15 to 49 Who Are Married or in a Union, 1970-2030. *Population Division Technical Paper* No 2013/2. New York, NY: United Nations, Department of Economic and Social Affairs, Population Division.

Kaufmann, E., Goujon, A. & Skirbekk, V. (2010). American political affiliation, 2003-43: A cohort component projection. *Population Studies*, 66(1), 53–67. https://doi.org/10.1080/00324728.2011.628047

KC, S., Barakat, B., Goujon, A., Skirbekk, V., Sanderson, W. C. & Lutz, W. (2010). Projection of populations by level of educational attainment, age, and sex for 120 countries for 2005-2050. *Demographic Research*, 22 (15), 383–472. https://doi.org/10.4054/DemRes.2010.22.15

KC, S. & Lutz, W. (2014). Demographic scenarios by age, sex and education corresponding to the SSP narratives. *Population and Environment*, 35 (3), 243–260. https://doi.org/10.1007/s11111-014-0205-4

KC, S., Wurzer, M., Speringer, M. & Lutz, W. (2018). Future Population and Human Capital in Heterogeneous India. *Proceedings of the National Academy of Sciences*: e201722359.
https://doi.org/10.1073/pnas.1722359115

Keilman, N. W. (1985). Internal and external consistency in multidimensional population projection models. *Environment and Planning* 17(11), 1473–1498. https://doi.org/10.1068/a171473

Keyfitz, N. (1977). *Introduction to the Mathematics of Population*. Second Edition. New York: Addison-Wesley.

Keyfitz, N. (1985). *Applied Mathematical Demography, 2nd edition*. New York, NY: Springer.

Kirk, D. (1996). Demographic Transition Theory. *Population Studies*, 50(3), 361–387.
https://doi.org/10.1080/0032472031000149536

Kreyenfeld, M. & Konietzka, D. (2017). *Childlessness in Europe: Contexts, Causes, and Consequences*. Springer.

Leahy, E., Lyons, S. & Tol, R. S. J. (2010). An Estimate of the Number of Vegetarians in the World. *ESRI Working Paper* No 340. Dublin: Economic and Social Research Institute.

Li, N. & Lee, R. D. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, 42(3), 575–594. https://dx.doi.org/10.1353%2Fdem.2005.0021

Loichinger, E. & Marois, G. (2018). Education-specific labour force projections for EU-28 countries. *In:* W. Lutz, A. Goujon, S. KC, M. Stonawski & N. Stilianakis (Eds.), *Demographic and Human Capital Scenarios for the 21st Century: 2018 assessment for 201 countries*, pp. 44–51. Luxembourg: Publications Office of the European Union. http://dx.doi.org/10.2760/835878

Lutz, W. (1994). *Population-Development-Environment: Understanding their Interactions in Mauritius*. Heidelberg: Springer-Verlag.

Lutz, W. (2013). Demographic Metabolism: A Predictive Theory of Socio-economic Change. *IIASA Research Report* (Reprint) RP-13-001. Laxenburg, Austria: International Institute for Applied Systems Analysis.
http://pure.iiasa.ac.at/10756

Lutz, W. & Skirbekk, V. (2014). How education drives demography and knowledge inform projections. *In:* W. Lutz, W. P. Butz & S. KC (Eds.), *World Population & Human Capital in the Twenty-first Century*, pp. 14–38. UK: Oxford University Press.

Lutz, W., Butz, W. P. & KC, S. (Eds.) (2014). *World Population & Human Capital in the Twenty-first Century*. UK: Oxford University Press.

Lutz, W., Crespo Cuaresma, J., Sanderson, W. C. (2008). The demography of educational attainment and economic growth. *Science*, 319 (5866), 1047–1048. https://doi.org/10.1126/science.1151753

Lutz, W. & Goujon, A. (2001). The World's Changing Human Capital Stock: Multi-State Population Projections by Educational Attainment. *Population and Development Review*, 27(2), 323–339.
https://www.jstor.org/stable/2695213

Lutz, W., Goujon, A., KC, S., Stonawski, M. & Stilianakis, N. (Eds.) (2018). *Demographic and Human Capital Scenarios for the 21st Century: 2018 assessment for 201 countries*. Luxembourg: Publications Office of the European Union. http://dx.doi.org/10.2760/835878

Lutz, W., Goujon, A. & Doblhammer-Reiter, G. (1998). Demographic Dimensions in Forecasting: Adding Education to Age and Sex. *Population and Development Review*, 24, Supplement: Frontiers of Population Forecasting (1998), 42–58. https://doi.org/10.2307/2808050

Lutz, W. & KC, S. (2010). Dimensions of global population projections: what do we know about future population trends and structures? *Philosophical Transactions of the Royal Society*, 365, 2779–2791. https://doi.org/10.1098/rstb.2010.0133

Manuel, M., Desai, H., Samman, E. & Evans, M. (2018). *Financing the end of extreme poverty Report*. London: Overseas Development Institute.

Margolis, R. (2016). The Changing Demography of Grandparenthood. *Journal of Marriage and Family*, 78, 610–622. https://doi.org/10.1111/jomf.12286

Marois, G., Sabourin, P. & Bélanger, A. (2019). How reducing differentials in education and labor force participation could lessen workforce decline in the EU-28. *Demographic Research*, 41(article 6), 125–160. https://dx.doi.org/10.4054/DemRes.2019.41.6

Moreira, P. A. & Padrão, P. D. (2004). Educational and economic determinants of food intake in Portuguese adults: a cross-sectional survey. *BMC Public Health*, 4(58). https://doi.org/10.1186/1471-2458-4-58.

Murray, G. R. & Mulvaney, M. K. (2012). Parenting Styles, Socialization, and the Transmission of Political Ideology and Partisanship. *Politics & Policy*, 40(6), 1106–1130. https://doi.org/10.1111/j.1747-1346.2012.00395.x

Ng, S. H. & Deng, F. (2017). Language and Power. *Oxford Research Encyclopedia*, Communication (oxfordre.com/communication). Oxford: University Press USA, 2019.

Notestein, F. (1945). Population: The long view. *In:* T. W. Schultz (Ed.), *Food for the World, pp. 36-57*. Chicago: University of Chicago Press.

O'Neill, B. C., Balk, D., Brickman, M. & Ezra, M. (2001). A Guide to Global Population Projections. *Demographic Research*, 4(8), 203–288. https://doi.org/10.4054/DemRes.2001.4.8

Petty, W. (1984). On the Causes and Consequences of Urban Growth. *Population and Development Review*, 10(1), 127–133. https://doi.org/10.2307/1973169

Pew Research Center (2015). *The Future of World Religions: Population Growth Projections, 2010-2050*. Washington, DC: Pew Research Center. https://www.pewforum.org/2015/04/02/religious-projections-2010-2050/

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F. S.,.. & Foley, J. (2009). Planetary boundaries: exploring the safe operating space for humanity. *Ecology and Society* 14(2), 32. http://www.ecologyandsociety.org/vol14/iss2/art32/

Rogers, A. [Ed] (1981). Advances in multiregional demography. *IIASA Research Report* 81-006. Laxenburg, Austria: International Institute for Applied Systems Analysis. http://pure.iiasa.ac.at/id/eprint/1556/1/RR-81-006.pdf

Rogers, A. & Land, K. (1982). *Multidimensional mathematical demography*. London: Academic Press.

Rooduijn, M. (2018). What unites the voter bases of populist parties? Comparing the electorates of 15 populist parties. *European Political Science Review*, 10(3): 351–368. https://doi.org/10.1017/S1755773917000145

Sabourin, P. & Bélanger, A. (2015). La dynamique des substitutions linguistiques au Canada. *Population*, 70(4) : 727–757. https://doi.org/10.3917/popu.1504.0771

Sandström, V., Valin, H., Krisztin, T., Havlík, P., Herrero, M. & Kastner, T. (2018). The role of trade in the greenhouse gas footprints of EU diets. *Global Food Security*, 19 (December 2018), 48–55. https://doi.org/10.1016/j.gfs.2018.08.007

Ortman, J. M. & Shin, H. B., (2011). Language Projections: 2010 to 2020. Presented at the Federal Forecasters Conference, Washington, DC, April 21, 2011. https://www.census.gov/content/dam/Census/library/working-papers/2011/demo/2011-Shin-Ortman.pdf

Springmann, M., Godfray, H. C. J., Rayner, M. & Scarborough, P. (2016). Analysis and valuation of the health and climate change cobenefits of dietary change. *Proceedings of the National Academy of Sciences*, 113 (15), 4146–4151. https://doi.org/10.1073/pnas.1523119113

Springmann, M., Wiebe, K., Mason-D'Croz, D., Sulser, T. B., Rayner, M. & Scarborough, P. (2018). Health and nutritional aspects of sustainable diet strategies and their association with environmental impacts: a global modelling analysis with country-level detail. *The Lancet*, 2(10), PE451-E461. https://doi.org/10.1016/S2542-5196(18)30206-7

Striessnig, E. & Lutz, W. (2016a). Demographic Strengthening of European Identity. *Population and Development Review*, 42(2), 305–311. https://dx.doi.org/10.1111%2Fj.1728-4457.2016.00133.x

Striessnig, E. & Lutz, W. (2016b). Demographic Metabolism at Work. *IIASA Working Paper* WP-16-001. Laxenburg, Austria: International Institute for Applied Systems Analysis. http://pure.iiasa.ac.at/id/eprint/12385/

UNDP (2014). *Human Development Report 2014. Sustaining Human Progress: Reducing Vulnerabilities and Building Resilience*. New York, NY: United Nations Development Programme. http://hdr.undp.org/sites/default/files/hdr14-report-en-1.pdf

United Nations (2015). *The World's Women 2015: Trends and Statistics*. New York: United Nations, Department of Economic and Social Affairs, Statistics Division. https://unstats.un.org/unsd/gender/worldswomen.html

United Nations, Department of Economic and Social Affairs, Population Division (2019). *World Population Prospects 2019*, Online Edition. https://population.un.org/wpp/

United Nations, Department of Economic and Social Affairs, Population Division (2018). *World Urbanization Prospects*: The 2018 Revision, Online Edition.

Whelpton, P. K. (1928). Population of the United States, 1925 to 1975. *American Journal of Sociology*, 34, 457–473. https://www.jstor.org/stable/2765600

WIC – Wittgenstein Centre for Demography and Global Human Capital (2018). Wittgenstein Centre Data Explorer Version 2.0 (Beta). http://www.wittgensteincentre.org/dataexplorer

Wils, A. B. (1996). PDE - Cape Verde: A Systems Study of Population, Development, and Environment. *IIASA Working Paper* WP-96-009. Laxenburg, Austria: International Institute for Applied Systems Analysis. http://pure.iiasa.ac.at/id/eprint/5016/1/WP-96-009.pdf

Wokes, F., Badenoch, J. & Sinclair, H. M. (1955). Human dietary deficiency of vitamin B12. *American Journal of Clinical Nutrition*, 3, 375–382. https://doi.org/10.1093/ajcn/3.5.375

Wunsch, G. J, & Termote, M. G. (1978). *Introduction to Demographic Analysis Principles and Methods*. New York, NY: Plenum.

Yousif, H. M., Goujon, A. & Lutz, W. (1996). Future Population and Education Trends in the Countries of North Africa. *IIASA Research Report* RR-96-011. Laxenburg, Austria: International Institute for Applied Systems Analysis. http://pure.iiasa.ac.at/id/eprint/4766/1/RR-96-011.pdf

Cash Accumulation by Non-Financial Corporations: New Evidence of the Role of Hedging Needs and Lower Financing Costs in France

Marie-Baïanne Khder* and Simon Ray**

Abstract – In this paper, we study the sources of the accumulation of cash by non-financial corporations in France. We notably explore cost-based explanations by proposing a firm-specific measure of the cost of carrying cash that depends on both the firms' short-term financing costs and the share of interest-bearing assets among liquid financial assets. Our analysis suggests that at least one fourth of the rise in the cash ratios between 2011 and 2016 is explained by the decreasing trend in the cost of carrying cash. When factoring in the additional impact of macroeconomic developments, our cost-based explanation accounts for up to 40% of the increase in cash holdings. We also identify a novel important determinant of the level of cash holdings: firms hold cash to seize future investment opportunities when they occur, irrespective of the financing conditions that will then prevail. Our results suggest that firms' cash hoarding to avoid foregone investment opportunities in downturns is an active economic stabilizer.

JEL Classification: G31
Keyword: financing frictions, investment, cash savings, debt capacity

**DG Trésor et CREST, Insee at the time of writing (marie-baianne.khder@dgtresor.gouv.fr) ; ** DG Trésor at the time of writing*

Acknowledgements – We thank A. Duquerroy, D. Blanchet, S. Roux, O. Simon, G. Lalanne, T. Tressel, C. Rousset and two anonymous referees for very helpful comments. We are also grateful to all participants to ECB's Workshop "France: structural challenges and reform" and the internal Insee seminar (DEE). We remain responsible for errors and omissions.

Received in July 2019, accepted in February 2020.

Citation: Khder, M.-B. & Ray, S. (2020). Cash Accumulation by Non-Financial Corporations: New Evidence of the Role of Hedging Needs and Lower Financing Costs in France. *Economie et Statistique / Economics and Statistics*, 520-521, 103–124. https://doi.org/10.24187/ecostat.2020.520d.2035

The recent sharp increase in financial liquid assets held by non-financial corporations (hereafter NFCs) has received considerable attention among policy makers, bankers and researchers in the field of corporate finance. This trend, though observed across a wide range of countries, has mainly been studied in the US. In this paper, we present new evidence on the recent increase in the share of financial liquid assets, and especially of cash holdings, in firms' balance sheets based on French firm-level data and we explore the determinants of cash hoarding.

The increase in French firms' cash holdings has been concomitant with a sustained rise in corporate debt, raising questions regarding the role that cash buffers could play to mitigate the risks associated with rising corporate debt (Khder & Rousset, 2017). Firms' recent cash accumulation is then inherently related to the issue of financial stability, and should also be linked to the transmission of monetary policy. Large cash buffers are likely to introduce a wedge, at least over the short term, between the funding cost of new projects and the level of the interest rates, potentially hampering the transmission of monetary policy. Corporate cash holdings also significantly affect the dynamic structure of banks' liabilities. This illustrates some of the first order macroeconomic and macro-financial consequences of corporates' cash management decisions. Despite the relevance of these questions, the economic literature has arguably not fully explored the determinants of corporates' cash holding. This paper intends to contribute to fill this gap.

Several explanations have been proposed for this shift in corporate cash holding based on the trade-off between the costs and benefits of cash from the perspective of shareholder wealth maximization, and empirical evidence has been provided mainly for the United States. With respect to the benefits, Bates *et al*. (2009), Boileau & Moyen (2016) and Bates *et al*. (2018) suggest that the volatility of corporate cash flows has increased over time, exacerbating firms' hedging needs and fostering precautionary savings, thus making cash holdings all the more valuable. Opler *et al*. (1999), Bates *et al*. (2009), Falato *et al.* (2013), Brown & Petersen (2013), Begenau & Palazzo (2017) and Adler *et al*. (2019) find that the surge in research and development (R&D) expenditure and intangible assets alters firms' ability to access external funding because these assets are relatively less pledgeable, therefore increasing the benefits derived from holding liquid financial assets. On the other hand, Azar *et al*. (2016) argue that the cost of carrying cash has shrunk.

In this paper, we study the evolution of cash holding in France since 2010 and document stylised facts on the dynamics of the cash level of French NFCs. Using firm level data, we explore the respective roles of original measures of costs and benefits associated with cash holdings. We examine the cost-based explanation for rising cash holdings using a new firm-level measure of the opportunity cost of carrying cash (that relies on the differences in firms' external financing costs and in firms' returns on short-term assets). In addition, we identify the role of the timing of investment opportunities on cash accumulation. Some firms choose to hold cash to hedge against the risk of foregoing a profitable investment opportunity because of low cash flows or tightened access to external finance at the time the investment opportunity occurs. We explore this explanation with an original methodology relying on sectoral local heterogeneity of the impact of the business cycle on firms' bankruptcy. We conduct the analysis over the period 2010-2016 on a rich dataset of firms' financial accounts merged with information on the capital linkages between social entities enabling to study the relevant aggregates at the group level.

Our analysis shows that cost-based explanation is the key to understanding the recently observed cash hoarding behaviours, in line with Azar *et al.* (2016). We document a semi-elasticity of the cash-to-asset ratio to the cost of carry of roughly 1.02. The average cost of carry in our database has shifted from 3.9% in 2011[1] to 2.3% in 2016. With our estimates, we explain up to 40% of the recent dynamics of the ratio of cash holdings to total assets (hereafter cash-to-asset ratio or cash ratio) by merely considering the change in the cost of carry that results from the aggregate fall in the cost of short-term financing.[2] When controlling for macroeconomic developments, the additional decrease in the cost of carrying cash at the firm level explains one fourth of the increase in cash holdings. We also document the significant role on cash levels of the hedging need against foregone investment opportunities, suggesting that firms' cash hoarding to avoid foregone investment opportunities in downturns is an important economic stabilizer.

1. Data on the cost of carry are available only since 2011.

2. Alternatively, if we consider the estimate (of Table 5, column 7) where firm-level observations are weighted by total asset size in the regression, and compare the evolution in the weighted mean of the cost of carry and of the cash ratio, we find that cost-based explanation explain 32% of the increase in cash holdings over 2010-2016. Weighting by total asset size – which is the denominator of our cash ratio- ensures extrapolation of our micro-level estimates to account for macro-level trajectory.

The remaining of the paper is organised as follows. The next section reviews the literature. The following section presents the data and the main descriptive statistics on French firms' cash holding. We then expose our empirical strategy in Section 3 and the results and interpretation of regression analyses in Section 4, before concluding.

1. Literature Review

Corporate cash holdings result from a trade-off between the costs and benefits of cash, as largely corroborated by the existing empirical literature: management that maximizes shareholder wealth should choose the level of firm' cash holdings such that their marginal benefit equals their marginal cost.

Let's first consider the costs of holding cash. Holding cash is costly because the spread between the marginal cost of external financing and the return on deposits or short-term financial investment is usually positive. Recent contributions argue that cost-based explanations are crucial to understand observed trends in corporate cash holdings. Azar *et al*. (2016) find that variations in the cost of carry, that is the cost of financing a dollar of liquid assets, net of the benefits derived from short-term financial investments, explain much of the secular increase in cash holdings since 1980 in the US. They also provide evidence of the preponderance of the cost-based explanation to cash accumulation in the five largest European economies and in Japan, based this time exclusively on national accounts data. Another source of costs associated with cash holding is the twofold tax disadvantages (Opler *et al.,* 1999), the income derived from liquid assets is taxed first at the corporate level as it increases the corporate income tax base and then, as for other assets, when income is distributed to shareholders because of income tax. Besides, the deductibility of interest payment may be capped; hence, an additional euro of debt invested in financial liquid asset can increase the corporate income tax base even when financing costs exceed the financial profits. However, because of the stability of the marginal corporate income tax rate in France over the period studied, tax-related explanation is unlikely to account for the recent dynamics of corporate cash holdings.[3]

As for the benefits, cash buffers enable firms to protect themselves against adverse cash flow shocks that could force them to liquidate assets or raise external funding at unfavourable conditions (hedging need against illiquidity and failure risk) and to finance investments regardless of the cost or the access to external financing (hedging need against foregone investment opportunities).[4] Indeed, as originally proposed by Keynes (1936), the main advantage of a liquid balance sheet is that it allows firms to undertake valuable projects when they arise irrespective of when external finance is cheap. Balance sheet liquidity is therefore all the more important that there exist frictions in the access to external financing. If a firm anticipates being financially constrained, its need to hedge against foregone investment opportunities is higher, as well as its optimal level of cash holding.

Linked to these two hedging motives, the literature emphasizes the impact of cash flow volatility on cash accumulation. Han & Qiu (2007) provide a theoretical foundation to this relationship when firms face financial constraints. Bates *et al*. (2009) or Boileau & Moyen (2016) identify the increase in cash flow volatility (Campbell *et al*., 2001 and Dichev & Tang, 2008 document this stylised fact) as one of the main factors explaining US firms' cash accumulation in the years 2000. To investigate the hedging need against foregone investment opportunities, some studies explore the effect of the correlation between cash flows and investment opportunities on cash hoarding. Acharya *et al*. (2007) develop a model predicting that financially constrained firms with high hedging needs – against foregone opportunities – have a strong propensity to save cash out of cash flows. In contrast, constrained firms with low hedging needs systematically channel cash flows towards debt reduction, as opposed to cash savings. They find strong empirical support in that sense. A key challenge to identify this mechanism is to measure the correlation between cash flows and investment opportunities: the apparent correlation between a firm's cash flows and investment spending is not relevant because the two are endogenously related when the firm is financially constrained. Acharya *et al*. (2007) consider two alternative measures of investment opportunities based on industry-level proxies. Since expenditures in R&D track growth opportunities, they first look at the correlation between a firm's cash flow from current operations and

3. *In the US, tax can also affect the level of cash because tax scheme may refrain multinational firms from repatriating cash from affiliates (Foley* et al.*, 2007) but such a channel arguably does not apply to France.*
4. *Opler* et al. *(1999) call the former channel the "transaction cost motive" and the latter the "precautionary motive", with reference to Keynes (1936). However, the meaning associated with these two terms has evolved in the literature. Therefore, we choose to refer henceforth to the former channel as "hedging need against illiquidity and failure risk" and to the latter as "hedging need against foregone investment opportunities".*

its industry-level median R&D expenditures to proxy the correlation between the firm's availability of internal funds and its unconstrained demand for investment. Their second measure consists in the correlation between firm-level cash flow and industry-level market demand, the latter being computed as the median three-year ahead sales growth rate in the firm's industry. However, these measures are arguably affected by the same financial constraints that prevent from merely using the observed correlations between cash flow and investment. In this paper, we assess an alternative sectoral local proxy that aims at capturing the impact of the correlation between cash flows and investment opportunities on cash accumulation.

Other factors have been put forward in the literature to explain firms' level of cash and its recent trend, notably R&D expenditure and the share of intangible capital in presence of financial frictions[5] (Opler *et al*., 1999; Bates *et al*., 2009; Begenau & Palazzo, 2017, who document sample-selection effects resulting from a shift toward less profitable "R&D-firms" that typically initially exhibit higher cash ratios going public; Falato *et al.,* 2013 or Adler *et al*., 2019) or information frictions (Jensen, 1986) – even if Opler *et al*. (1999), Bates *et al*. (2009), Kalcheva & Lins (2007) do not find significant evidence of the influence of principal-agent problem on cash hoarding.

2. Data and Descriptive Statistics on Cash and Liquidity Accumulation by French Corporations

2.1. Data Sources and Consolidation Method

We use administrative data provided by the French Statistical Institute (Insee) and covering the period 2010-2016. We work at the group level. Indeed, although they file stand-alone accounts, legal units are not necessarily autonomous in their economic decision-making process because of the numerous financial and customer-supplier and operational linkages they are involved in as parts of corporate groups. Consolidating the accounts of legal units is therefore necessary for the quality of the analysis. As shown in Picart (2003), productive activities and financial management activities are likely to be allocated to distinct legal units belonging to the same corporate group. Cash flows are often transferred from legal units involved in production to legal units incorporated for financial management purposes. Some assets, such as real estate, are also often borne by separate legal units with specific legal status (*Sociétés Civiles Immobilières* for instance), which are in turn more likely to bear the related debt liabilities (Insee, 2019). The existence of intra-group cash transfers, as evidenced by Locorotondo *et al*. (2014), provides support to our assumption that corporate financial policy decisions, in particular regarding cash management, are made at group level, echoing previous research (Lamont, 1997). The level of consolidation matters because it substantially affects the usual financial ratios (Deroyon, 2015) and, as expected, the variation in cash ratios is much larger when computed at the legal unit level than after consolidation. This excess variability at the legal unit level reflects measurement errors due to intra-group reallocation rather than the decision to hoard cash made by groups on economic grounds in the face of variations in financing conditions, warranting consolidation. Finally, consolidation also fosters the comparability with international studies based on datasets such as ORBIS, Compustat (collecting consolidated accounts released by groups in annual reports).

Because our preferred statistical unit is the group, we consolidate financial statements from the "raw" database of legal units ESANE (*Élaboration des statistiques annuelles d'entreprise*). A group is a set of legal units linked by capital ownership, that are identified using the *LIaisons FInancieres* (LIFI) database, an administrative dataset providing information about the ownership and nationality of the parent company of firms located in France.[6] Based on the raw accounts of legal units, we create for each corporate group a new statistical observation, the "pseudo-group". For each group, the financial statement of the corresponding pseudo-group is calculated from those legal units belonging to the core of the group (i.e. owned at more than 50% by the parent company,[7] and therefore controlled by the group).[8] Our final database is

5. The idea that financially constrained firms have significantly lower payout ratios follows from Fazzari et al. *(1988) and Fama & French (2002). Alternative approaches to distinguish groups of financially constrained and unconstrained firms merely rely on the firms' size, as in Erickson & Whited (2000). Fama & French (2002) and Frank & Goyal (2003) also associate firm size with the degree of external financing frictions. Other measures of financial constraints are based on credit rating, and notably on the fact of having a credit rating or not (e.g. Whited, 1992; Lemmon & Zender, 2001).*
6. The survey is exhaustive on the set of firms that employ more than 500 employees, that generate more than 60 million euros in revenues or that hold more than 1.2 million euros of shares, but is completed by data coming from Bureau Van Dijk (Diane-Amadeus data set) to cover the whole universe of French business groups.
7. The parent company (tête de groupe) *is the legal unit that owns the majority of other legal units without being in turn owned in majority by them.*
8. In a previous version of the paper, the financial statement of a pseudo-group was computed from all the legal units constituting the group, pro rata to parent company's ownership rate in the legal unit. The main regression results were unchanged with this alternative consolidation methodology.

composed, unless otherwise stated, of three types of statistical units: *(i)* pseudo-groups based on consolidation restatements of core legal units, *(ii)* legal units related to business groups but not controlled by them, henceforth called legal units loosely related to groups and *(iii)* independent legal units, not belonging to any group. Legal units at the core of business groups are excluded from the final database once consolidated (to avoid double-counting with pseudo-groups). Our consolidation approach however suffers from some shortcomings: our automatic consolidation is less accurate than consolidation carried out by Insee, which is based on additional data and ongoing discussion with the accountants of larger groups (this does not extend yet to all firms). The coverage of the LIFI database varies over 2010-2016, introducing potential additional measurement errors. Our data on balance sheet items and profit and loss statements cover exclusively the French perimeter of groups, therefore leading to measurement errors for highly internationalized groups. Further details are provided in Appendix 1.

The raw financial data in ESANE come from the balance sheet information collected from firms' tax forms, which covers the universe of French legal units, excluding the financial and agricultural sectors as a rule. In this study, we focus on firms in the private sector and restrict the analysis to the normal tax regime (called BRN for *Bénéfice réel normal*) because it covers most of the total amount of liquid financial assets. Throughout the study, the sector is defined at the group level for pseudo-groups (based on LIFI database which provides a group-level sector), and at the legal unit level for legal units that are independent or loosely related to business groups. Finally, the location of a consolidated pseudo-group is defined as the region where the largest number of legal units belonging to its core are located.

2.2. Examining the Sample

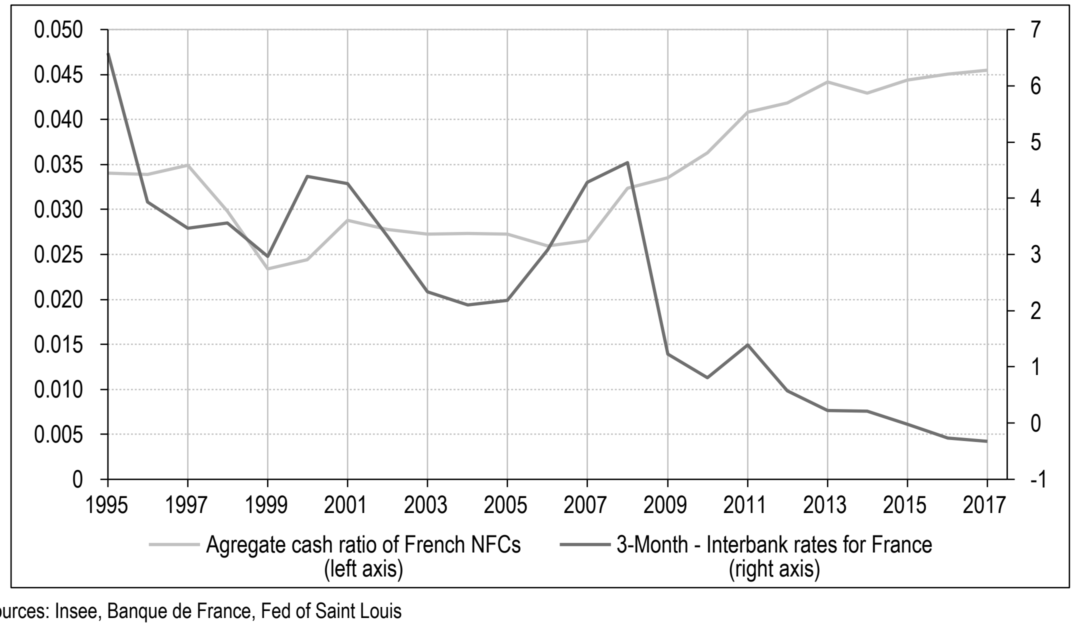
In this section, we present further evidence on the levels and dynamics of corporate cash holdings and financial liquid assets.

We observe a negative relationship over a longer period between the average cash-to-asset ratio of NFCs and the level of short-term interest rates as measured by the 3-month interbank rates for France[9] (Figure I). In both series, we notice a clear concomitant break, in opposite directions, since the financial crisis.

Analysis at group level allows to track the dynamics of the distribution of cash and financial liquid asset ratios. We observe upwards trends for most of the moments of these distribution, suggesting an overall rightward shift of the distribution (Figure II). Nevertheless, we notice a more pronounced growth of the third quartile suggesting an increase in the concentration of cash holdings. The median cash ratio increased by 3.6 percentage points (pp)[10] between 2010 and 2016 to reach 13.9% in 2016. The rise of

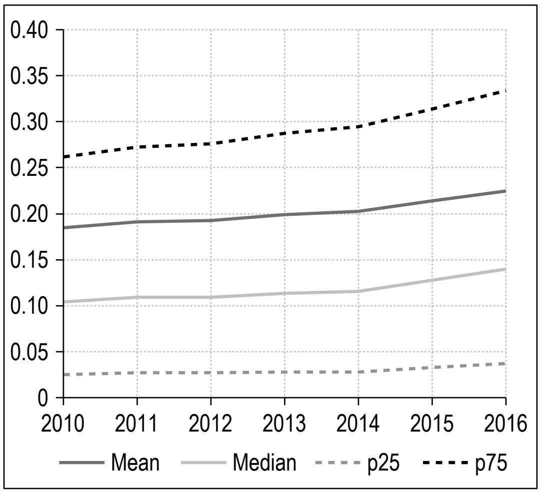
9. Data on NFCs cash in the national financial account are available from 1995 onwards.
10. percentage of total asset.

Figure I – **Aggregate cash to assets ratio and short-term interest rates in % – national accounts**



Sources: Insee, Banque de France, Fed of Saint Louis

Figure II – **Moments of the cash to assets ratios**



Sources: Insee (Esane/LIFI); authors' calculations.

the median liquid financial asset ratio is less pronounced: only 1.5 pp over the studied period (see Figure A3-I in Appendix 3). Indeed, in the context of very low, or even negative, interest rates, the return on the non-cash interest-bearing financial liquid assets held by NFCs, both short-term debt securities and money market funds (MMFs), has declined. In this environment, firms substituted MMFs for cash; however, firms still have overall increased their holdings of financial liquid assets.

The upward trend is also pervasive across sectors (Figure III). The median levels of financial liquid asset ratios are nevertheless heterogeneous across activities, and the highest in sectors such as professional, scientific and technical services, information and communication and other services. These sectors have also experienced the highest increase in their financial liquid asset ratio, in line with the results of the literature (e.g., Opler *et al.*, 1999; Bates *et al.*, 2009) linking cash holding to intangible assets and financial frictions.

Small firms (10 to 249 employees) and micro firms (less than 10 employees) tend to hold more cash as a percentage of total asset than larger firms (Figure IV). Size is a major determinant of financial liquid asset holdings. Both the relative position across size categories and the level of the ratios that we document are comparable to what Bates *et al.* (2009) evidenced in the US.

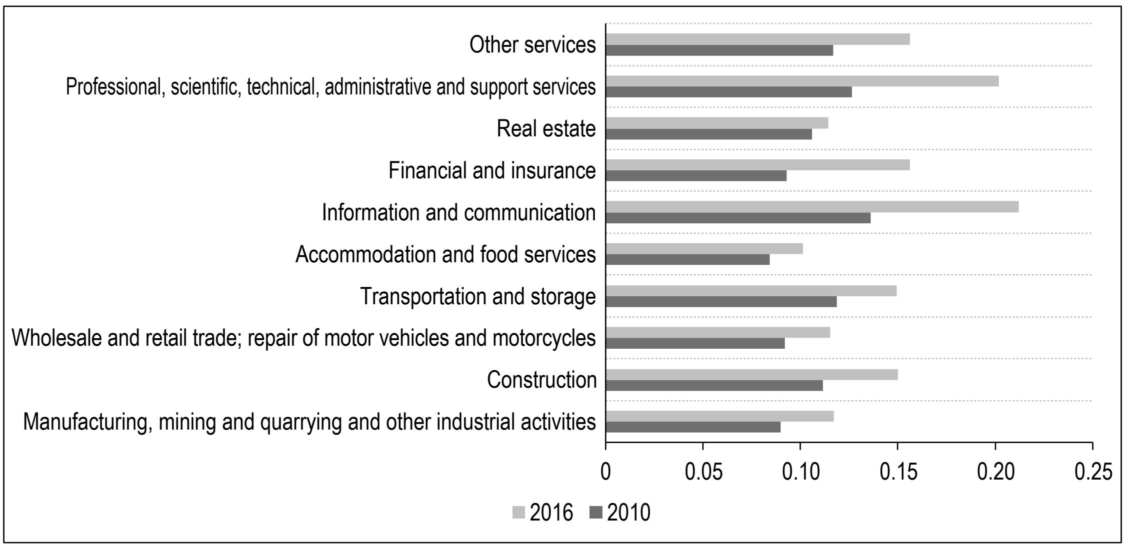
Alternatively, we use exclusively for this paragraph another sample including core legal units fully controlled by a group, but without consolidation restatements, legal units loosely related to a group and independent legal units. With this sample, the median levels of cash asset ratios in the three subsets tend to follow similar upward trends (Figure V); independent legal units exhibit much higher cash ratios than their peers belonging to a group.

3. Empirical Strategy

3.1. The Cost of Carry at the Firm Level

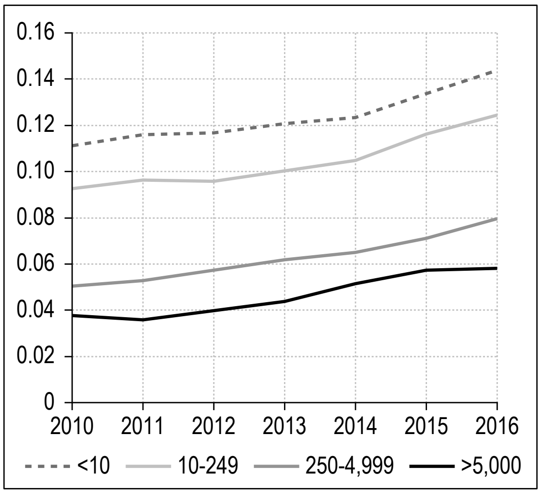
The cost of carrying cash and financial liquid assets corresponds to the difference between the cost of an extra euro of external funding and the return of this extra euro when it is held as liquid financial assets, part of it being deposited in cash accounts or invested in

Figure III – **Median of the cash to assets ratios by industry**



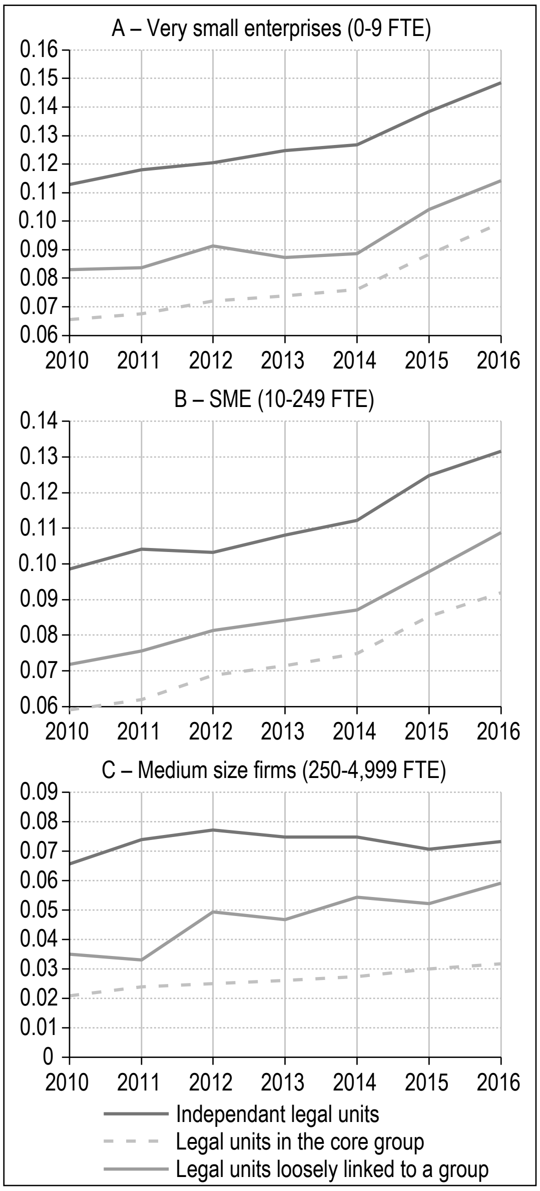
Sources: Insee (Esane/LIFI); authors' calculations.

Figure IV – **Median of the cash to assets ratios by firms' size**



Note: <10 refers to firms (independent, loosely related to groups or after consolidation) with below 10 full time equivalent employees.
Sources: Insee (Esane/LIFI); authors' calculations.

Figure V – **Median of the cash to assets ratios by size and status**



Sources: Insee (Esane/LIFI), authors' calculations.

short-term interest-bearing financial assets. The cost of carrying cash varies across firms because, on the one hand, the cost of external financing depends on the firm's creditworthiness and, on the other hand, the return on liquid financial asset may differ according to the allocation between interest-bearing and non-interest-bearing assets. Azar *et al*. (2016) explore exclusively this second source of variation to derive a firm-specific cost of carry. With respect to the first source of variability of external financing across firms, they assume in their empirical analysis that the cost of external financing is equal to the 3-month T-Bill rate for all firms. Their assumption that "because cash is a risk-free investment, the cost of capital should correspond to the risk-free rate" does not hold since the cost of capital depends on the overall financial soundness of the firm, and consequently the perceived counterparty risk.

Unlike Azar *et al*. (2016), we exploit both sources of variation across firms of the cost of carry. We therefore introduce a novel proxy of the cost of external financing that a firm is likely to face based on the assessment of its credit risk. This proxy relies on moments of the cost of short-term debt reported by the Banque de France.[11] For each year, we assess the firms' creditworthiness[12] through the Altman Z''-score (Altman, 1983[13]) (see Appendix 1). Altman's Z''-score predicts the probability of business failure, which influences the cost at which a firm can raise additional debt. As exemplified by the 2019 Global Financial Stability Report published by the IMF, it is used, among other tools, by practitioners to gauge a firm's credit strength. We match firm observations and the

11. The moments of the distribution of the annual interest rates on new debt contract - i.e. p5, p25, p50, p75 – are computed by the Banque de France based on their database MContran.

12. We choose to impute a cost of short-term external financing based on credit worthiness rather than using the apparent cost of debt (defined as the ratio of interest payments to outstanding debt) because (i)*, credit constrained firms do not by definition report debt in the tax file: this would bias our sample towards non-financially constrained firms;* (ii) *the apparent cost of debt indicates the average price of one unit of debt, whereas we conceptually focus on the marginal cost of one extra unit of debt.*

13. Altman's Z''-score (1983) consists of a linear combination of EBITDA/ total assets, working capital requirements/total assets, accumulated retained earnings/total assets, and equity at historical cost/total assets. This score is designed to assess the probability of failures of private and publicly listed manufacturing and non-manufacturing companies, but was estimated in 1983 on a limited sample of companies. Nevertheless, Altman et al. *(2017) rejects, on the basis of ORBIS dataset composed of roughly 2.7 million observations from European firms, the hypothesis of an obsolescence of the parameters estimated in Altman (1983)'s Z''-score in terms of classification performance.*

annual cost of short-term debt by merging the percentile of the creditworthiness distribution and the percentile of the cost of debt.[14] With respect to the second source of firm-level variation in the cost of carry pertaining to the return on financial liquid assets, following Azar *et al.* (2016), we use the firm-level share of short-term interest-bearing securities in financial liquid assets in the first year of observation (to alleviate the endogeneity concerns linked to the cost of carry). We assume that short-term investment securities generate an annual return equal to the average annual performance of the money market funds, as published by the Banque de France. The firm-level formula of the cost of carry (CoC) then writes:

$$CoC_{it} = Cost\ of\ short\ term\ debt_{pct\,i,t} - share_{i,t_0}\, perfMMF_t \quad (1)$$

Moments of the distribution of firm-specific cost of carry are reported in Table 1. The cost of carrying cash sharply declined between 2011 and 2016, with the mean (the median) value of the CoC decreasing by 1.44 pp (respectively 1.26 pp).[15]

Using the Z''-score as a source of identifying variation in our regressions raises some endogeneity issues. For instance, investors could interpret high cash holdings as a sign of financial soundness, enabling the firm to contract new loans (reverse causality). The increase in leverage would be translated via the Z''-score into a lower decrease in the cost of external funding. This could bias downward (in absolute terms) our estimate of the elasticity of the cash ratio to the cost of carry. However, we first decide to include lagged values of the cost of carry, to mitigate as much as possible endogeneity concerns. Then, we use the percentiles of Z''-score, and not the Z''-score *per se*. This allows us to alleviate, though not totally discard, endogeneity concerns.

Alternatively, and as a robustness check, we use a cost of carry based exclusively on a firm-level measure of the cost of short-term debt. For this alternative indicator, we match firms with the moments of the distribution of the cost of short-term debt based on the SAFE score (Ferrando *et al.*, 2015) rather than the Z''-score. The SAFE-score aims at measuring the extent of financial constraints faced by firms. It consists of the weighted sum of a firm's financial ratios.[16] The weights are estimated based on the financial constraints, as reported in the survey on the Access to Finance of Enterprises (SAFE) from the ECB, on a sample of micro, small, medium-sized and large European firms from 2010 to 2013. Our preferred measure remains the cost of carry based on the Z''-score, notably because the endogeneity concerns might be more acute for the SAFE-score due to the inclusion of cash-to-asset ratio in its definition.

3.2. A New Measure of the Correlation Between Cash Flows and Investment Opportunities

As mentioned in the literature review, theoretical contributions have highlighted that the correlation between cash flows and investment opportunities explains the accumulation of cash by firms, while underlining the difficulty to identify empirically this correlation due to endogeneity concerns (Acharya *et al.*, 2007). Investment opportunities may arise in a state of the world where a firm has low positive cash flows and is subsequently more likely to face financial constraints. In this case, the firm highly values cash holdings, because they would allow seizing an investment opportunity in the future despite low earnings or tightened access to external financing. Firms that are already constrained and not profitable in good times do not have the opportunity to hoard cash to seize future investment opportunities. However, firms whose financial situation is sound enough in good times, but which anticipate a tightening of their access to external financing in bad times, might hoard cash to hedge against foregone investment opportunities.

One key driver of a negative correlation between cash flows and investment opportunities results from assets or firms being sold at distressed

14. We choose not to control for the sector while ranking firm according to their percentile of creditworthiness. Indeed, a sector as a whole could be characterized as a below-average creditworthiness.
15. The weighted (by total asset) mean of the cost of carry decreases by 1.1 pp.
16. Namely the debt/total asset ratio, the paid interest/retained earnings ratio, the profit margin ratio, the tangible asset/total asset ratio, the cash holdings/total asset ratio, and the logarithm of total asset.

Table 1 – **Moments of the distribution of the cost of carry**

Cost of carry	Number of observations	Number of values	Mean	sd	q10	q25	Median	q75	q90
2011	578,061	138,949	3.86	1.58	2.16	2.75	3.43	4.42	6.65
2016	639,551	162,883	2.42	1.30	0.92	1.51	2.17	2.80	4.71

Sources: Insee (Esane/LIFI), Banque de France; authors' calculations.

prices due to fire sales. As suggested by Shleifer & Vishny (2011, p. 30), "a fire sale is essentially a forced sale of an asset at a dislocated price. The asset sale is forced in the sense that the seller cannot pay creditors without selling assets. The price is dislocated because the highest potential bidders are typically involved in a similar activity as the seller and are therefore themselves indebted and cannot borrow more to buy the asset. […] Assets are then bought by non-specialists who, knowing that they have less expertise with the assets in question, are only willing to buy at valuations that are much lower." The frequency and the magnitude of such an event vary across industries and, to some extent, when the relevant secondary market for assets is at least partially local. The intuition is the following: during slowdowns, pressure on firms to "fire sell" their assets, the most extreme pressure being business failure, increases. In sectors and regions where this pressure is the highest, the relative value of holding cash is the greatest, because firms that managed to accumulate enough can make the most of the more numerous fire sales of assets. We do not have proper direct measures of assets prices on the secondary market that would capture sectoral and local specificities with respect to fire sales. The effect of economic growth on the frequency of business defaults at the sector-region level provides a relevant proxy of the exposure to investment opportunities at distressed prices. We then recover the sector-region elasticities of business failures to the economic cycle by estimating the following regression equation:

$$Default_{s,r,t} = \beta_{s,r}\Delta g_t + \alpha_{s,r} + \delta_t + \epsilon_{s,r,t} \qquad (2)$$

where $Default_{s,r,t}$ is the number of business failures[17] registered in sector s, region r and year t normalized by the number of firms operating[18] in sector s, region r at year t, $\beta_{s,r}$ captures the sector-region sensitivity of defaults to the economic cycle, $\alpha_{s,r}$ are sector-region fixed effects capturing the average local sectoral level of default and δ_t are year fixed effects. Estimation runs from 1994 to 2009.[19] Δg_t refers to GDP growth in year t. Sectors are broadly defined because of the structure of the data on defaults produced by the Banque de France (first level of the French classification of activities), and agriculture and non-profit sectors are excluded. Regions are the new French regions after the territorial reform of 2014. The coefficients of interest in equation (2) are the $\beta_{s,r}$, which correspond to the sector-region

17. Business failures at the sectoral local level are disseminated by the Banque de France based on the FIBEN data. The FIBEN database is truncated to the left (sales > 75 000 euros), the number of failures might therefore be underestimated.

18. The number of firms operating in each sector and each region over 1994-2009 is computed using Insee's SIRENE databases.

19. The $\beta_{s,r}$ elasticities are estimated prior to the main regressions (which aim at explaining the level and dynamics of cash ratios) to mitigate endogeneity concerns.

Table 2 – **$\beta_{s,r}$ sectoral local elasticities of business failures to the economic cycle**

Region/Sector	Manufacturing	Construction	Wholesale and retail trade	Transportation and storage	Accommodation and food service	Information and communication	Finance and insurance	Real estate	Administrative and support services	Other services	Mean	Sd.
Île-de-France	-1.33	-1.17	-1.06	-0.83	-1.09	-1.06	-1.37	-1.09	-0.94	-0.71	*-1.1*	*0.2*
Centre-Val de Loire	-0.98	-0.83	-0.78	-0.81	-0.88	-0.78	-2.04	-0.99	-0.68	-0.68	*-0.9*	*0.4*
Bourgogne Franche-Comté	-0.86	-0.80	-0.80	-0.71	-0.82	-0.78	-1.42	-0.80	-0.74	-0.62	*-0.8*	*0.2*
Normandie	-0.96	-1.05	-0.85	-0.81	-0.85	-0.61	-1.87	-1.02	-0.81	-0.67	*-1.0*	*0.4*
Hauts-de-France	-1.09	-0.98	-0.79	-0.75	-0.98	-0.79	-1.72	-0.79	-0.75	-0.65	*-0.9*	*0.3*
Grand Est	-0.93	-1.09	-0.85	-0.84	-0.80	-0.65	-2.13	-0.85	-0.93	-0.69	*-1.0*	*0.4*
Pays de la Loire	-0.88	-0.83	-0.88	-0.76	-0.81	-0.78	-1.38	-0.75	-0.75	-0.68	*-0.8*	*0.2*
Bretagne	-1.11	-0.73	-0.80	-0.75	-0.94	-0.99	-1.68	-0.79	-0.74	-0.71	*-0.9*	*0.3*
Nouvelle-Aquitaine	-0.93	-0.84	-0.84	-0.74	-0.98	-0.89	-1.52	-0.83	-0.74	-0.66	*-0.9*	*0.2*
Occitanie	-1.06	-0.96	-0.85	-0.68	-0.85	-0.96	-1.73	-0.84	-0.80	-0.67	*-0.9*	*0.3*
Auvergne Rhône-Alpes	-0.93	-0.93	-0.82	-0.68	-0.86	-0.73	-1.58	-0.95	-0.77	-0.67	*-0.9*	*0.3*
Provence Alpes Côte d'Azur	-1.12	-1.15	-0.89	-0.71	-0.98	-1.00	-1.95	-0.96	-0.88	-0.65	*-1.0*	*0.4*
Corse	-1.08	-1.35	-0.77	-0.81	-1.56	n.a.	0.45	-0.68	-0.24	-0.34	*-0.7*	*0.6*
Mean	-1.0	-1.0	-0.8	-0.8	-1.0	-0.8	-1.5	-0.9	-0.8	-0.6		
Sd.	0.1	0.2	0.1	0.1	0.2	0.1	0.6	0.1	0.2	0.1		

Sources: Insee (Esane/LIFI), Banque de France; authors' calculations.

elasticities of business failures to the economic cycle. The associated estimates vary widely across sectors and across regions. The most negative $\beta_{s,r}$ are found for sectors such as financial and insurance services, construction, or manufacturing. Conversely, business failures in services appear less sensitive to the cycle. Our estimates also reveal some heterogeneity across regions within sectors. Table 2 reports the estimated value of $\beta_{s,r}$.

The sector-region elasticities of business failures to the cycle, $\beta_{s,r}$, might partially capture hedging needs against illiquidity and failure risk, in addition to the hedging needs against foregone investment opportunities that we would like to isolate. To purge as much as possible our elasticities of the illiquidity hedging needs, we introduce $\alpha_{s,r}$ sector-region fixed effects (not interacted with GDP growth) in equation (2).[20] They are more likely to capture hedging needs against illiquidity and failure than the $\beta_{s,r}$. Indeed, we assume that firms assess their own probability of failure based on the average sectoral local number of business defaults (captured by the $\alpha_{s,r}$) rather than on the sensitivity of business failures to the cycle (captured by the $\beta_{s,r}$). We will also provide additional robustness tests below to disentangle those two channels.

As a robustness check, we present an alternative measure of hedging need against foregone investment opportunities, which relies on the gross amount of business failures rather than on the normalized business failures, therefore changing the dependent variable in equation (2).

In addition to these two variables of interest, we build control variables identified in the literature as important determinant of firms' cash and liquid financial holdings. The list of variables used in our regression and information regarding the way they are built are presented in Table 3.

4. Estimation and interpretation

4.1. Panel Regression with Firm Fixed Effects

We first estimate a model where yearly firm-level cash ratios[21] are regressed on firm fixed effects, which capture the role of observed and unobserved time invariant firms' characteristics on cash holdings, and on a set of time-varying observable characteristics. Including firm fixed effects enables to capture the effect of the change in the cost of carry at the firm level on the change in the cash ratio. Year fixed effects are included as robustness checks to control

20. In a previous version of the article, from equation (2) were included in our main regressions (with the cash ratio as dependent variable), with a positive and significant influence on cash hoarding, consistently with the hedging against illiquidity motive. For the sake of clarity, we excluded this control variable, with no influence on other estimation coefficients.

21. Regressions with the financial liquid assets/ total asset ratio as dependent variable convey consistent conclusions.

Table 3 – **Variables**

Variable	Description
Cash / assets ratio (narrow definition)	Cash (CF in the tax files) divided by total assets consolidated at the group level
Cash / assets ratio[1] (extended definition)	Financial liquid assets (CF + CD in the tax files) divided by total assets consolidated at the group level
Cost of carry (CoC)	Firm level cost of short-term funding (based on Z"-score) minus revenues derived from short term financial assets (defined by eq (1))
Cost of short-term debt	Firm level cost of short-term funding (based on the SAFE-score)
Z"- score	Z-score based on net working capital/asset, EBIT/asset, retained earnings/asset and equity/asset as defined in Altman (1983) – percentiles are built based on annual distributions
Financial debt / assets	Consolidated financial debt (DS+DT+DU in the tax files) divided by total assets consolidated at the group level – Intragroup financial debt are fully excluded
Pay-out ratio	Dividend paid by the parent company divided by the consolidated after-tax results
ln(Assets)	Log of the total assets consolidated at the group level
Earnings / assets	Retained earnings divided by the total assets consolidated at the group level
Share of tangible	Tangible assets divided by the total assets consolidated at the group level
SD(EBIT)	Firm-level standard deviation of the level of the Earnings Before Interests and Taxes over the observation period, measuring the volatility of cash flows. Divided by 1000 for presentation purpose.
Hedging needs	The correlation between investment opportunities and the firm's cash flows. It is computed as the correlation between the median industry-level R&D spending and the firm's earnings.
$\beta_{s,r}$	Sector-region elasticities of business failures to the cycle as defined and estimated in Eq (2) and reported in Table 2 – the more negative the $\beta_{s,r}$, the more sensitive to the economic cycle the firms in sector-region, i.e. the more numerous business failures in case of economic downturns

[1] Also called financial liquid assets to total assets ratio.
Notes: For further details, see Appendix 1, and statistics for these variables in Appendix 3.

Table 4 – **Model with firm fixed effects (dependent variable: cash to assets ratio)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	(narrow)	(extended)	(narrow)	(extended)	(narrow)	(extended)	(narrow)	(extended)
Cost of carry(-1)	-0.0102*** (0.0001)	-0.0100*** (0.0001)	-0.0059*** (0.0007)	-0.0078*** (0.0010)			-0.0092*** (0.0001)	-0.0084*** (0.0001)
Cost of short-term debt(-1)					-0.0084*** (0.0001)	-0.0076*** (0.0001)		
Net working capital/ Assets	-0.0676*** (0.0011)	-0.0761*** (0.0011)	-0.0664*** (0.0041)	-0.0754*** (0.0036)	-0.0767*** (0.0003)	-0.0863*** (0.0003)	-0.0559*** (0.0003)	-0.0546*** (0.0004)
Financial debt/ Asset(-1)	-0.0020*** (0.0007)	-0.0023*** (0.0007)	-0.0012 (0.0010)	-0.0019 (0.0013)	0.0004 (0.0003)	-0.0014*** (0.0003)	0.0098*** (0.0002)	0.0020*** (0.0002)
Earnings/Asset	0.0412*** (0.0010)	0.0444*** (0.0010)	0.0408*** (0.0028)	0.0442*** (0.0034)	0.0343*** (0.0004)	0.0370*** (0.0004)	0.0038*** (0.0005)	-0.0458*** (0.0007)
ln(Asset)	-0.0475*** (0.0008)	-0.0416*** (0.0008)	-0.0457*** (0.0013)	-0.0407*** (0.0017)	-0.0360*** (0.0003)	-0.0316*** (0.0003)	-0.0060*** (0.0001)	-0.0020*** (0.0002)
Payout ratio	-0.0065*** (0.0003)	-0.0042*** (0.0003)	-0.0043*** (0.0012)	-0.0032** (0.0015)	-0.0065*** (0.0002)	-0.0042*** (0.0002)	-0.0023*** (0.0002)	-0.0023*** (0.0002)
Fixed Effect	Firm-FE	Firm-FE	Firm-FE & Year-FE	Firm-FE & Year-FE	Firm-FE	Firm-FE	Firm-FE	Firm-FE
Clustering	Firm	Firm	Firm and year	Firm and year	None	None	None	None
Weight	None	None	None	None	None	None	Asset size	Asset size
Sample	Full	Full	Full	Full	Full	Full	Full	Full
Observations	2,473,753	2,473,753	2,473,753	2,473,753	2,124,721	2,124,721	2,473,753	2,473,753
R2	0.82	0.86	0.82	0.86	0.83	0.87	0.88	0.91
Adjusted R2	0.74	0.80	0.74	0.80	0.75	0.81	0.83	0.86

Notes: *p<0.1; **p<0.05; ***p<0.01. Variables definitions are given in Table 3. Robust standard errors are reported in parenthesis.
Sources: Insee (Esane/LIFI); authors' calculations.

for exogenous year-specific factors that could contribute to the average increase in the cash ratio between 2010 and 2016. The main results are reported in Table 4. Our measures of CoC is included with a one-year lag to mitigate endogeneity issues.[22]

The estimates of the coefficient associated with our CoC are negative and highly statistically significant across all specifications. Based on the estimate in our first specification (Column 1), the change in the mean value of the CoC between 2011 and 2016 explains a change of 1.5 pp in the level of the cash ratio, which is over 40% of the mean increase in the cash ratios over the period. This indicates that cost-based explanations are of paramount importance to understand the recent cash accumulation by firms. The significant effect of the CoC is robust to the inclusion of year fixed effects (column 3): over 25% of the mean increase in the cash ratios is explained by the decrease in the cost of carrying cash. The effect of the CoC is then identified only from within-firm changes in the risk premium, i.e. to put it differently, our result is not identified by the overall downward trend in the cost of funding resulting from expansionary monetary policy over the period of interest because of the year fixed effects. From column 5, we infer that a decrease in our alternative measure of short-term external financing cost also significantly increases cash hoarding, although this alternative measure ignores the share of interest-bearing financial liquid assets. The negative and significant impact of the CoC on cash ratios holds and is quantitatively similar when we weight firm-level observations by total assets while running the regression (column 7). This enables to draw conclusions on the “macro-evolution” of cash holdings: based on the reported estimate, the change in the mean value of the CoC between 2011 and 2016 explains again roughly 40% of the mean increase in the cash ratios over the period. Our results are also robust to an extended definition of the cash ratio, when cash holdings in the numerator also include marketable securities and own shares beyond mere cash accounts and bank deposits (columns 2, 4, 6, 8). Finally, our results are robust to balancing the panel, and even reinforced (see Appendix 2, Table A2-3 columns 3 and 4).

22. An increase in cash holdings contemporaneously with the diminution of external cost of funding might reflect the accounting rather than the economic phenomenon at play. When the firm's cost of carry decreases, it is more likely to raise financial debt, which gives rise to financial resources recorded on the asset side of the balance sheet as cash before this additional resource is used for investment. On the contrary, if cash holdings increase following a previous decrease in the cost of carry, the firm made the economic decision to keep as cash the additional financial resources it has raised, without a assigning those resources to specific investment in the short-term. Finally, endogeneity concerns are only partially alleviated: our model does not allow for fully-fledged causal identification.

We introduce in these regressions a set of time varying variables: an increase in cash holdings go hand in hand with a decrease in net (of cash balances) working capital and with an increase in firm's annual earnings. We include control variables to capture the level of the financial frictions faced by firms. The regression results largely corroborate the negative relationship existing between the ease of access to external financing at the extensive margin (i.e. the degree of financial constraints faced by the firm) and cash accumulation. Firms that exhibit higher payout ratios (which often reflects low financial constraints as documented above) tend to have lower cash ratios. Hadlock & Pierce (2010) document that the higher the size of total asset, the less likely the firm is to be financially constrained. Consequently, the statistically significant negative impact of firm size on cash holdings indicates that firms having easier access to external financing due to lower financial constraints hold less cash.

Cash holdings react differently to an increase in indebtedness across size categories. The positive and significant correlation between indebtedness and cash hoarding once weighting firm-level observations by total assets size suggests that easier access to external funding has fuelled cash accumulation (see below Table 5 columns 7, 8). This result corroborates and extends the findings of Khder & Rousset (2017). Besides, consolidation is required for the adequacy of the analysis: for large consolidated pseudo-groups, while a positive and significant correlation between indebtedness and cash accumulation is evidenced, similar regressions run on their constitutive legal units does not reveal such a correlation (see Appendix 2, Table A2-1 column 3). For large and mid-sized firms, an increase in the lagged ratio of financial debt to total assets is positively correlated with an increase in the cash ratio while they are negatively and significantly correlated for SMEs (Appendix 2, Table A2-2). For large firms, the effect of the CoC on cash accumulation is no longer significantly negative. This could be attributed to the nature of the Z''-score at the heart of the CoC, which aims at predicting business failures, and is therefore a more accurate proxy of external cost of financing for small firms than for large firms.

As a final robustness check, we estimate this model on subsamples composed of pseudo-groups only, of independent legal units only, and on legal units belonging to corporate groups (Appendix 2, Table A2-1). The main take-away from this analysis is that cash hoarding behaviours are affected by changes in the CoC for independent legal units and for pseudo-groups, across all class size roughly. Among very small and small firms, the impact of the decrease in the cost of carry seems to be higher for independent legal units than for legal units belonging to a corporate group.

This first set of regressions presented in Table 4 highlights the key role of cost-based explanations in recent trends observed at the macroeconomic level. The next sub-section further explores these dimensions.

4.2. Panel Regression with Sectoral Fixed Effects

Because of the firm fixed effects, we are not able in the first set of regressions to estimate the coefficients associated with sector-region elasticities of business failures to the economic cycle $\beta_{s,r}$, since they do not vary over time. Thus, we run similar regressions replacing firm fixed effects by sector and region fixed effects.[23] Sector and region fixed effects are necessary to control for sectoral time-invariant and regional time-invariant features that could otherwise bias the estimate on our sectoral regional elasticities. The new regression model also enables to estimate the effect of other firm-level time-invariant characteristics put forward in the literature as important determinants of the level of cash holdings that could not be identified with the previous regressions such as the volatility of earnings (Bates et al., 2009).

Estimation results with sector-year fixed effects[24] are in line with those obtained when we only exploit within-firm variations (Table 5). We find a statistically significant negative effect on cash hoarding of the different measures of the cost of carry. Regarding the time-invariant characteristics introduced in these regressions, we estimate statistically significant coefficient with the expected signs. We find that firms characterized by more volatile EBIT over the observed period hold higher levels of cash.

The estimates associated with our novel measure $\beta_{s,r}$ are negative and statistically significant. This result suggests that the higher the sectoral local elasticities of business failures to the cycle (which means more negative $\beta_{s,r}$), or in other words the more numerous the investment opportunities at fire sale prices are, the more firms hoard cash. The estimated effects of this

*23. Sectors are defined at the NAF (*Nomenclature d'activités françaises*) 5-digit level and regions are defined according to the 2014 territorial reform.*
24. Our results are also robust to the inclusion of sector-year fixed effects instead of sector fixed effects, to capture sector-level time-varying shocks such as sectoral demand shocks.

variable are noticeable. Based on the estimated negative coefficient associated with the $\beta_{s,r}$ in Table 5 (column 1), in Ile-de-France, 5 pp of the difference in cash ratios between firms in the business services sector and firms in the manufacturing sector are explained by the elasticities $\beta_{s,r}$. Incidentally, the effect is robust to the inclusion of year fixed effects (column 3).

In the robustness test (column 2), our alternative measure of investment opportunities $\beta_{s,r}$ is as expected significantly and negatively correlated with cash ratio, and with final effect[25] on cash ratio of the same order of magnitude that in our baseline specification (column 1). As another robustness check, we run a regression where firm-level observations are weighted by the size of total assets (column 4): the coefficient on the elasticities $\beta_{s,r}$ is significant and negative, and 10 times higher than in the similar though unweighted regression (column 1). This shows that sectoral-regional elasticities of business failures are more important for larger firms and are therefore likely to matter at a macroeconomic level. This also hints at the higher likelihood for larger firms to hedge against foregone investment opportunities by hoarding cash.

At this stage, we cannot discard a selection bias in our empirical framework. Our finding that, in sectors and regions where business defaults are very sensitive to the cycle, firms tend to hoard more cash could stem from a bias in our sample towards surviving firms: the most financially distressed firms, with less cash holdings, might have failed and exited the sample. Correcting for this selection bias (Heckman, 1979) requires a valid instrument for the probability of exiting that does not influence the volume of cash and debt. We do not have such an instrument. However, running our regression model with a balanced or quasi-balanced[26] sample shows that the effect on cash hoarding of sectoral regional elasticities is robust, and even reinforced, when

25. The estimate is facially substantially lower than for the baseline measure of investment opportunities in column 1, but this alternative measure of investment opportunities is not normalized by the number of firms in a given sector-region and thus higher in absolute terms.
26. The quasi-balanced panel includes firms that are present in the database all years except one.

Table 5 – **Model with sectoral fixed effects (dependent variable: cash to assets ratio, narrow definition)**

	(1)	(2)	(3)	(4)
Cost of Carry (-1)	-0.0102***	-0.0102***	-0.0083***	-0.0058***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Net Working Capital/Assets	-0.0418***	-0.0418***	-0.0399***	-0.0975***
	(0.0002)	(0.0002)	(0.0002)	(0.0003)
Financial debt/Assets (-1)	-0.0093***	-0.0093***	-0.0091***	0.0058***
	(0.0002)	(0.0002)	(0.0002)	(0.0001)
sd(EBIT)	0.0253***	0.0253***	0.0251***	0.0001***
	(0.0007)	(0.0007)	(0.0007)	(0.00002)
Earnings/Assets	0.0298***	0.0298***	0.0306***	0.0935***
	(0.0004)	(0.0004)	(0.0004)	(0.0008)
ln(Asset)	-0.0323***	-0.0323***	-0.0318***	-0.0087***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
β_{s,r} (baseline)	-0.0051**		-0.0052**	-0.0407***
	(0.0022)		(0.0022)	(0.0015)
β_{s,r} (alternative)		-0.00004***		
		(0.00001)		
Payout ratio	0.0323***	0.0323***	0.0340***	0.0262***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Tangible assets/Asset	-0.2605***	-0.2605***	-0.2605***	-0.2050***
	(0.0009)	(0.0009)	(0.0009)	(0.0006)
Fixed Effects	Sector-FE & Region-FE	Sector-FE & Region-FE	Sector-FE & Region-FE & Year-FE	Sector-FE & Region-FE
Clustering	None	None	None	None
Weight	None	None	None	Asset size
Sample	Full	Full	Full	Full
Observations	2,151,394	2,151,573	2,151,394	2,151,394
R2	0.15	0.15	0.15	0.29
Adjusted R2	0.15	0.15	0.15	0.29

Notes: *p<0.1; **p<0.05; ***p<0.01. Variables definitions are given in Table 3.
Sources: Insee (Esane/LIFI); authors' calculations.

considering surviving firms only: the selection bias does not seem to matter to the first order (Appendix 2, Table A2-3 columns 1 and 2).

4.3. Disentangling the Various Mechanisms Captured by the Sectoral Local Elasticities $\beta_{s,r}$

With Table 6, we address the interpretation of the significant effect of our sectoral local elasticities $\beta_{s,r}$ of business defaults to the cycle on cash hoarding. The effect of the $\beta_{s,r}$ evidenced so far could capture two distinct channels:

- hedging needs against foregone investment opportunities: some firms, which can afford to accumulate cash in good times, do so because they anticipate in bad times to be financially constrained or to have low earnings and they would like to seize the investment opportunities that could occur in their sector and region because of fire sales of assets during slowdowns;
- hedging needs against illiquidity and failure risk: some firms may hoard cash to avoid defaults and failures, regardless of investment opportunities.

We argue that the sectoral regional elasticities $\beta_{s,r}$ mostly capture the hedging needs against foregone investment opportunities. To disentangle the contribution of those two channels, we first observe that they have different implications depending on asset specificity in a sector. On the one hand, if the hedging needs against foregone investment opportunities prevail, the more specific assets are to the sector, the more advantage a firm with high cash holdings can draw from asset fire sales of its competitors within the same sector and region. Firms would therefore value cash holdings more in sectors featuring high asset specificity. On the other hand, if the hedging needs against illiquidity prevail, the impact of our sectoral local elasticities should only marginally depend on asset specificity. We proxy the degree of asset specificity to a given sector by the ease with which the assets used in the sector can be redeployed across other sectors following Kim & Kung (2016). We distinguish here by a dummy "high asset specificity" the sectors in which assets are the least easily redeployable across other sectors based on Kim & Kung (2016).[27] The effect of the elasticities $\beta_{s,r}$ on cash holdings is significantly higher (roughly 10 times higher) in sectors where assets are the most sector-specific (column 1): a high sectoral local elasticity of business failures to the cycle triggers cash hoarding almost exclusively in sectors where assets are sector-specific. This suggests the $\beta_{s,r}$ capture first-order hedging needs against foregone investment opportunities, rather than against illiquidity risk.

In columns 2 and 3, we contrast the effect of hedging needs against foregone investment opportunities with that of the real option channel (Pindyck, 1991; Bloom, 2009), which states that, when an investment is irreversible, the firm postpones investment in the face of uncertainty, and values more cash because of the embedded option to invest in the future it provides. In column 2, we show that policy uncertainty, as measured by the Economic Policy Uncertainty index (Baker *et al.* 2016), significantly and positively affects cash holdings, in line with the real option theory.[28] The effect of the $\beta_{s,r}$ elasticities on cash is however robust (in significance and in order of magnitude, cf. Table 5, column 1) to the inclusion of the economic policy uncertainty index. Besides, we find that higher economic policy uncertainty does not lead to higher cash hoarding in sectors with highest degree of assets specificity (cf. column 3; the interaction term is even significantly negative). [29] The take-away is that the greater effect on cash hoarding of $\beta_{s,r}$ elasticities in sectors with a high degree of asset specificity arguably cannot be attributed to the real option channel, and thus that our $\beta_{s,r}$ elasticities primarily measure hedging needs against foregone investment opportunities, rather than the real option channel.

In column 6, we interact the $\beta_{s,r}$ elasticities with the quintile of size. As a reminder, (total asset) size is often considered as a proxy of financial constraints (along with age, cf. Hadlock & Pierce, 2010): the larger the firm, the easier the access to external financing. We find that the effect of the $\beta_{s,r}$ elasticities on cash hoarding is larger for the fourth and the fifth quintiles of asset size, namely the 40% largest firms. For those large firms, the significant negative coefficient associated with the elasticities mostly reflect the hedging needs against foregone investment opportunities. On the contrary, for the lowest quintiles of size, firms are small and, when they operate in sectors and regions where business failures are highly sensitive to the economic cycle (i.e. when the $\beta_{s,r}$

27. Namely, the high asset specificity dummy turns to 1 for the 15 out of 53 industries with the least redeployable assets based on Kim & Kung's index, see Kim & Kung (2016) Table 1. [Textile mills, Semiconductor and electronic component manufacturing, Plastics and rubber products manufacturing, etc.]

28. This effect is not more pronounced for industries with higher investment irreversibility.

29. This is somehow at odds with the real option theory that suggests that firms would hoard more cash in sectors where investment is highly specific (and then more likely to be more irreversible) and in times of higher uncertainty.

Table 6 – **Model with sectoral fixed effects, further investigation (dependent variable: cash to assets ratio, narrow definition)**

	(1)	(2)	(3)	(4)	(5)
Cost of Carry (-1)	-0.0083***	-0.0100***	-0.0100***	-0.0087***	-0.0065***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Net Working Capital/Asset	-0.0399***	-0.0415***	-0.0415***	-0.0402***	-0.0389***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0003)
Financial debt/Asset (-1)	-0.0091***	-0.0093***	-0.0093***	-0.0093***	-0.0084***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
sd(EBIT)	0.0251***	0.0253***	0.0253***	0.0235***	0.0232***
	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0008)
Earnings/Asset	0.0306***	0.0300***	0.0300***	0.0371***	0.0275***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
ln(Asset)	-0.0318***	-0.0322***	-0.0322***	-0.0447***	-0.0308***
	(0.0001)	(0.0001)	(0.0001)	(0.0003)	(0.0001)
β_{s,r}	-0.0027	-0.0052**	-0.0027		
	(0.0023)	(0.0022)	(0.0023)		
Policy uncertainty		0.0001***	0.0001***		
		(0.000004)	(0.000004)		
Payout ratio	0.0340***	0.0321***	0.0321***	0.0344***	0.0340***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Tangible assets/Asset	-0.2606***	-0.2607***	-0.2607***	-0.2607***	-0.2639***
	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0010)
β_{s,r}*high asset specificity	-0.0477***		-0.0470***		
	(0.0077)		(0.0077)		
Policy uncertainty*high asset specificity			-0.00004**		
			(0.00002)		
β_{s,r}*size quintile 1				0.0160***	
				(0.0023)	
β_{s,r}*size quintile 2				0.0196***	
				(0.0023)	
β_{s,r}*size quintile 3				0.0043*	
				(0.0022)	
β_{s,r}*size quintile 4				-0.0197***	
				(0.0023)	
β_{s,r}*size quintile 5				-0.0556***	
				(0.0023)	
Hedging needs*size quintile 1					0.0012**
					(0.0006)
Hedging needs*size quintile 2					-0.0071***
					(0.0006)
Hedging needs*size quintile 3					-0.0099***
					(0.0006)
Hedging needs*size quintile 4					-0.0142***
					(0.0006)
Hedging needs*size quintile 5					-0.0224***
					(0.0007)
Fixed Effects	Sect-FE & Reg-FE & Year-FE	Sector-FE & Region-FE	Sector-FE & Region-FE	Sect-FE & Reg-FE & Year-FE	Sect-FE & Reg-FE & Year-FE
Clustering	None	None	None	None	None
Weight	None	None	None	None	None
Sample	Full	Full	Full	Full	Full
Observations	2,151,394	2,151,394	2,151,394	2,151,394	1,814,221
R2	0.15	0.15	0.15	0.16	0.15
Adjusted R2	0.15	0.15	0.15	0.16	0.15

Notes: *p<0.1; **p<0.05; ***p<0.01. Variables definitions are given in Table 3.
Sources: Insee (Esane/LIFI); authors' calculations.

elasticities are more negative), they tend to be more fragile (because they did not have the time to build cash buffers and their business environment is volatile). The primary objective of those smaller firms is to hedge against illiquidity and failure.

Finally, we include Acharya *et al*. (2007)'s measure of hedging needs against foregone investment opportunities in the regression (column 5). This alternative proxy consists of the correlation between investment opportunities and firms' cash flows. It is computed as the correlation between the median industry-level R&D spending[30] and firms' earnings (cf. literature review). First, and in line with Acharya *et al.* (2007)'s findings, we document that the lower the correlation between investment opportunities and cash flows (the hedging needs indicator is thus negative), the higher the cash ratio. This increase in the cash ratio is statistically significantly across almost all asset size categories. Second, the impact of hedging needs on cash hoarding monotonously increases with the size of assets: larger firms (which are typically less financially constrained) with higher hedging needs (i.e. a more negative hedging need indicator) tend to hoard more cash. This result, slightly different from Acharya *et al*.'s claim that only firms with higher financial constraints and high hedging needs choose to hoard cash rather than to reduce debt, hints at the fact that firms that can afford to carry cash (less financially constrained for instance) do so when they anticipate that investment opportunities might arise in times where the firms' cash flows may lag behind: this also provides evidence that the hedging motive against foregone investment opportunities plays a significant role to understand firms' cash hoarding behaviours.

* *
*

In this paper, we explore the sources of the cash accumulated by NFCs and the determinants of the sharp increase in the cash and liquid financial assets ratios recently observed in France.

We take advantage of firm-level variations in the cost of carrying cash derived from heterogeneous costs of short-term financing to document that this variable largely explains the recent trends. We also find robust evidence that financial constraints and hedging needs are key determinants of firm-level cash accumulation. Based on an original measure of the correlations between cash flows and investment opportunities, that are proxied by the local sectoral elasticities of business defaults to the economic cycle, we document that hedging needs against foregone investment opportunities explain the large difference in the levels of cash across regions and across sectors. Our results have important policy implication, notably with regard to financial stability. They suggest in particular that the current level of cash could be significantly altered in the event of a trend reversal in the cost paid by firms for short-term debts but also that firms' cash buffers are likely to dampen fire sales mechanisms in the upcoming crisis as firms seem to hoard cash in anticipation of the investment opportunities arising in the economic downturns. This result suggests that firms' cash hoarding is an active economic stabilizer. This question could be further explored in future research. ☐

30. We exploit the R&D survey by the French ministry of Higher education and Research. Industry is defined at the A88 level.

BIBLIOGRAPHY

Acharya, V. V., Almeida, H. & Campello, M. (2007). Is cash negative debt? A hedging perspective on corporate financial policies. *Journal of Financial Intermediation*, 16(4), 515–554.
https://doi.org/10.1016/j.jfi.2007.04.001

Adler, K., Ahn, M. J. & Dao, M. C. (2019). *Innovation and Corporate Cash Holdings in the Era of Globalization*. International Monetary Fund.
https://www.imf.org/en/Publications/WP/Issues/2019/01/18/Innovation-and-Corporate-Cash-Holdings-in-the-Era-of-Globalization-46494

Almeida, H., Campello, M. & Weisbach, M. S. (2004). The cash flow sensitivity of cash. *The Journal of Finance*, 59(4), 1777–1804. https://doi.org/10.1111/j.1540-6261.2004.00679.x

Altman, E.I. (1983). *Corporate Financial Distress: A Complete Guide to Predicting, Avoiding, and Dealing With Bankruptcy*. Wiley Interscience, John Wiley and Sons

Antoun de Almeida, L., Batini N., Tressel T. & Voigts S. (2018). France: selected issues. IMF Country Report No. 18/244. https://www.imf.org/en/Publications/CR/Issues/2018/07/26/France-Selected-Issues-46127

Azar, J. A., Kagy, J. F. & Schmalz, M. C. (2016). Can changes in the cost of carry explain the dynamics of corporate "cash" holdings? *The Review of Financial Studies*, 29(8), 2194–2240. https://doi.org/10.1093/rfs/hhw021

Baker, S. R., Bloom, N. & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4), 1593–1636. https://doi.org/10.1093/qje/qjw024

Bates, T. W., Kahle, K. M. & Stulz, R. M. (2009). Why do US firms hold so much more cash than they used to? *The Journal of Finance*, 64(5), 1985–2021. https://doi.org/10.1111/j.1540-6261.2009.01492.x

Bates, T. W., Chang, C. H. & Chi, J. D. (2018). Why has the value of cash increased over time? *Journal of Financial and Quantitative Analysis*, 53(2), 749–787. http://dx.doi.org/10.2139/ssrn.1975491

Begenau, J. & Palazzo, B. (2017). *Firm selection and corporate cash holdings*. *NBER Working Paper* N° w23249. https://doi.org/10.1016/j.jfineco.2020.09.001

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, 77(3), 623–685. https://doi.org/10.3982/ECTA6248

Boileau, M. & Moyen, N. (2016). Corporate cash holdings and credit line usage. *International Economic Review*, 57(4), 1481–1506. https://doi.org/10.1111/iere.12205

Brown, J. R. & Petersen, B. C. (2011). Cash holdings and R&D smoothing. *Journal of Corporate Finance*, 17(3), 694–709. https://doi.org/10.1016/j.jcorpfin.2010.01.003

Campbell, J. Y., Lettau, M., Malkiel, B. G. & Xu, Y. (2001). Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. *The Journal of Finance*, 56(1), 1–43. https://doi.org/10.1111/0022-1082.00318

Deroyon, J. (2015). De nouvelles données issues du profilage des groupes : une part accrue de l'industrie, des entreprises plus performantes, mais une capacité à financer l'investissement et un endettement plus dégradés. *In:* Insee*, Les entreprises en France 2015.* https://www.insee.fr/fr/statistiques/1906511?sommaire=1906539

Dichev, I. D. & Tang, V. W. (2008). Matching and the changing properties of accounting earnings over the last 40 years. *The Accounting Review*, 83(6), 1425–1460. https://doi.org/10.2308/accr.2008.83.6.1425

Erickson, T. & Whited, T. M. (2000). Measurement error and the relationship between investment and q. *Journal of Political Economy*, 108(5), 1027–1057. https://www.journals.uchicago.edu/doi/abs/10.1086/317670

Falato, A., Kadyrzhanova, D. & Sim, J. (2013). Rising intangible capital, shrinking debt capacity, and the US corporate savings glut. Available at SSRN: http://dx.doi.org/10.2139/ssrn.3198030

Fama, E. F. & French, K. R. (2002). Testing trade-off and pecking order predictions about dividends and debt. *The Review of Financial Studies*, 15(1), 1–33. https://doi.org/10.1093/rfs/15.1.1

Faulkender, M. W., Hankins, K. W. & Petersen, M. A. (2017). Understanding the rise in corporate cash: Precautionary savings or foreign taxes. *The Review of Financial Studies*, 32(9), 3299–3334. https://doi.org/10.1093/rfs/hhz003

Fazzari, S. M., Hubbard, R. G. & Petersen, B. C. (1988). Financing Constraints and Corporate Investment. *Brookings Papers on Economic Activity*, 19(1), 141–206. https://www.brookings.edu/wp-content/uploads/1988/01/1988a_bpea_fazzari_hubbard_petersen_blinder_poterba.pdf

Ferrando, A., Blank, S., Neugebauer, K., Siedschlag, I., Iudice, M., Altomonte, C., ... & Meinen, P. (2015). Assessing the financial and financing conditions of firms in Europe: the financial module in CompNet. ECB Working Paper No. 1836. Available at SSRN: https://ssrn.com/abstract=2641070

Foley, C. F., Hartzell, J. C., Titman, S. & Twite, G. (2007). Why do firms hold so much cash? A tax-based explanation. *Journal of Financial Economics*, 86(3), 579–607. https://doi.org/10.1016/j.jfineco.2006.11.006

Hadlock, C. J. & Pierce, J. R. (2010). New evidence on measuring financial constraints: Moving beyond the KZ index. *The Review of Financial Studies*, 23(5), 1909–1940. https://doi.org/10.1093/rfs/hhq009

Han, S. & Qiu, J. (2007). Corporate precautionary cash holdings. *Journal of Corporate Finance*, 13(1), 43–57. https://doi.org/10.1016/j.jcorpfin.2006.05.002

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the Econometric Society,* 47(1), 153–161. https://doi.org/10.2307/1912352

IMF (2019). Global Financial Stability Report, October 2019: Lower for Longer. https://www.imf.org/en/Publications/GFSR/Issues/2019/10/11/Global-Financial-Stability-Report-October-2019-Lower-for-Longer-47080

Insee (2019). Productivité du travail et du capital : une mesure renouvelée au niveau de l'entreprise. *In:* Insee Références, *Les entreprises en France.* https://www.insee.fr/fr/statistiques/4255787?sommaire=4256020

Jensen, M. C. (1986). Agency costs of free cash flow, corporate finance, and takeovers. *The American Economic Review*, 76(2), 323–329. https://www.jstor.org/stable/1818789

Kalcheva, I. & Lins, K. V. (2007). International evidence on cash holdings and expected managerial agency problems. *The Review of Financial Studies*, 20(4), 1087–1112. https://doi.org/10.1093/rfs/hhm023

Keynes, J. M. (1936) 2007. *The General Theory of Employment, Interest, and Money.* Palgrave Macmillan.

Khder, M. B. & Rousset, C. (2017). Faut-il s'inquiéter de la hausse de l'endettement des entreprises en France? Insee*, Note de conjoncture décembre 2017.*
https://www.insee.fr/fr/statistiques/3292331?sommaire=3292415

Kim, H. & Kung, H. (2016). The asset redeployability channel: How uncertainty affects corporate investment. *The Review of Financial Studies*, 30(1), 245–280. https://doi.org/10.1093/rfs/hhv076

Lamont, O. (1997). Cash flow and investment: Evidence from internal capital markets. The *Journal of Finance*, 52(1), 83–109. https://doi.org/10.1111/j.1540-6261.1997.tb03809.x

Lamont, O., Polk, C. & Saaá-Requejo, J. (2001). Financial constraints and stock returns. *The Review of Financial Studies*, 14(2), 529–554. https://doi.org/10.1093/rfs/14.2.529

Lemmon, M. L. & Zender, J. F. (2010). Debt capacity and tests of capital structure theories. *Journal of Financial and Quantitative Analysis*, 45(5), 1161–1187. https://www.jstor.org/stable/27919560

Locorotondo, R., Dewaelheyns, N. & Van Hulle, C. (2014). Cash holdings and business group membership. *Journal of Business Research*, 67(3), 316–323. https://doi.org/10.1016/j.jbusres.2013.01.019

Pindyck, R. S. (1991). Irreversibility, Uncertainty, and Investment. *Journal of Economic Literature*, 29(3), 1110–1148. https://doi.org/10.3386/w3307

Pindyck, R. S. (2000). Irreversibilities and the timing of environmental policy. *Resource and Energy Economics*, 22(3), 233–259. https://doi.org/10.1016/S0928-7655(00)00033-6

Nikolov, B. & Whited, T. M. (2014). Agency conflicts and cash: Estimates from a dynamic model. *The Journal of Finance*, 69(5), 1883–1921. https://doi.org/10.1111/jofi.12183

Opler, T., Pinkowitz, L., Stulz, R. & Williamson, R. (1999). The determinants and implications of corporate cash holdings. *Journal of Financial Economics*, 52(1), 3–46. https://doi.org/10.1016/S0304-405X(99)00003-3

Shleifer, A. & Vishny, R. (2011). Fire sales in finance and macroeconomics. *Journal of Economic Perspectives*, 25(1), 29–48. https://doi.org/10.1257/jep.25.1.29

Whited, T. M. (1992). Debt, liquidity constraints, and corporate investment: Evidence from panel data. *The Journal of Finance*, 47(4), 1425–1460. https://doi.org/10.1111/j.1540-6261.1992.tb04664.x

APPENDIX 1

THE CONSOLIDATION METHOD

Controlling for Intra-Group Operations

A variable at the group-level is not necessarily the mere sum of the variables of its core legal units. Some variables can be directly summed across all the legal units within a group, because they do not include intra-group flows (respectively intra-group stocks) or because intra-group operations cancel out in summation over the group. Consolidation can thus be carried out directly on the values reported by the legal units for:

- Employment;
- Cash;
- Financial liquid assets other than cash;
- Tangible and intangible fixed capital, and investment;
- Earnings, defined here as earnings minus interests, taxes, depreciation, amortization, and dividends;
- Earnings Before Interest and Taxes (EBIT).

On the other hand, some variables include intra-group operations that do not cancel out when summed over the group, and that would thus lead to double-counting. An example is the outstanding financial debt amount. Intra-group indebtedness turns out to be critical given its recent rise, as evidenced by de Almeida *et al*. (2018). To consolidate non-cumulative variables at the group-level, our preferred approach is: *(i)* to build at the legal unit level a new variable cleaned of intra-group items; *(ii)* to carry out consolidation restatement on those “cleaned” variables to construct the pseudo-group’s variable. For non-cumulative variables, step *(i)* is performed as follows:

- Total asset: at the legal unit-level, we retrieve out of the total asset (minus depreciation) the intra-group loans, and financial participation in legal units and loans associated to financial participations;

- Financial debt: we restrict ourselves to a convertible bonds, other bonds and loans by credit institutions. This excludes all intra-group loans;

- Dividends: for a group, we keep only the dividend paid by the parent company, since the other legal units in the group do not own the parent company. The parent company therefore necessarily pays dividends to outside shareholders.

Altman (1983)’s Z”-score

$$Z''score = 3.25 + 6.56\ WorkingCapital + 3.26\ Earnings + 6.72\ EBIT + 1.05\ Equity$$

where *Working Capital* corresponds to Working capital / Total assets, *Earnings* to Cumulative retained earnings / Total assets, *EBIT* to EBIT / Total assets, and *Equity* to Book value of equity / Total assets.

Ferrando *et al.* (2015) SAFE-score

$$\text{SAFE-score} = -1.88 + 0.86\ Finlev + 0.28\ ipf + 0.51\ profitmargin - 0.21\ collateral - 1.21\ cashholdings - 0.05 \ln(TotalAssets)$$

where *Finlev* refers to Financial debt / Total assets, *ipf* (the index of financial pressure) to Interest payments / Earnings, *profitmargin* to EBIT / Sales, *collateral* to Fixed assets / Total assets and *cashholdings* to Cash holdings / Total assets.

Building the Sectoral Local Elasticities of Business Failures to the Cycle for Pseudo-Groups

For a group, we construct the sectoral local elasticities of business failures to the cycle as a weighted average of the sectoral local elasticities of all the constitutive legal units. The weights are computed as the lagged share of the legal unit in the assets of the group.

ADDITIONAL REGRESSIONS

Table A2-1 – **Model with firm fixed effects, by firm status (dependent variable: cash to assets ratio)**

	Pseudo-groups	Independent legal units	Legal units in a group
Cost of carry lag1 x <10	-0.0023***	-0.0064***	-0.0033***
	(0.0007)	(0.0007)	(0.0005)
Cost of carry lag1 x 10-249	-0.0035***	-0.0059***	-0.0023***
	(0.0007)	(0.0008)	(0.0008)
Cost of carry lag1 x 250-4,999	-0.0028***	-0.0033	0.0003
	(0.0010)	(0.0023)	(0.0020)
Cost of carry lag1 >5,000	-0.00001		0.0061
	(0.0034)		(0.0040)
NWC / asset	-0.0229***	-0.0774***	-0.0222***
	(0.0048)	(0.0052)	(0.0015)
Earnings / Asset	0.0182***	0.0428***	0.0075***
	(0.0042)	(0.0033)	(0.0004)
ln(Asset)	-0.0334***	-0.0442***	-0.0412***
	(0.0018)	(0.0016)	(0.0009)
Payout ratio	-0.0023	-0.0044***	-0.0019**
	(0.0018)	(0.0012)	(0.0009)
Financial debt / Asset x <10	-0.0052	-0.0046	0.0001
	(0.0066)	(0.0035)	(0.0004)
Financial debt / Asset x 10-249	-0.0062*	-0.0058	-0.0003
	(0.0035)	(0.0043)	(0.0006)
Financial debt / Asset x 50-4,999	0.0102	0.0515	0.0128**
	(0.0089)	(0.0472)	(0.0065)
Financial debt / Asset x >5,000	0.0526**		-0.0005
	(0.0221)		(0.0080)
Full sample	Group level	Indep. legal units	Legal units in groups
Firm-FE	Yes	Yes	Yes
Sector-FE	No	No	No
Year-FE	Yes	Yes	Yes
SE-Clustering	Firm+Year	Firm+Year	Firm+Year
Observations	276,405	2,038,952	1,393,598
R2	0.86	0.82	0.79
Adjusted R2	0.78	0.74	0.70

Notes: *p<0.1; **p<0.05; ***p<0.01. Variables definitions are given in Table 3.
Sources: Insee (Esane/LIFI); authors' calculations.

Table A2-2 – **Model with firm fixed effects, by firm size (dependent variable: cash to assets ratio, narrow definition)**

	Large firms	Mid-size firms	SMEs
Cost of Carry (-1)	-0.0049	-0.0093***	-0.0116***
	(0.0054)	(0.0015)	(0.0003)
Net Working Capital/Asset	-0.0095	-0.0212	-0.1193***
	(0.0237)	(0.0209)	(0.0057)
Financial debt/Asset (-1)	0.0176	0.0094	-0.0063***
	(0.0170)	(0.0105)	(0.0017)
Earnings/Asset	-0.0392	0.0209	0.0887***
	(0.0305)	(0.0216)	(0.0044)
ln(Asset)	-0.0208	-0.0051	-0.0001
	(0.0203)	(0.0090)	(0.0017)
Payout ratio	-0.0223	-0.0044	-0.0023***
	(0.0196)	(0.0065)	(0.0005)
Fixed Effects	Firm-FE	Firm-FE	Firm-FE
Clustering	Firm	Firm	Firm
Weight	None	None	None
Sample	> 5,000 FTE	250 FTE - 4,999 FTE	10 FTE - 249 FTE
Observations	553	1,209	541,628
R2	0.81	0.86	0.87
Adjusted R2	0.74	0.79	0.81

Notes: *p<0.1; **p<0.05; ***p<0.01. Variables definitions are given in Table 3.
Sources: Insee (Esane/LIFI); authors' calculations.

Table A2-3 – **Model with balanced sample and quasi-balanced sample (dependent variable: cash to assets ratio, narrow definition)**

	(1)	(2)	(3)	(4)
Cost of Carry (-1)	-0.0067***	-0.0062***	-0.0114***	-0.0125***
	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Net Working Capital/Asset	-0.0433***	-0.0478***	-0.0708***	-0.0767***
	(0.0003)	(0.0003)	(0.0013)	(0.0016)
Financial debt/Asset(-1)	-0.0076***	-0.0078***	-0.0012*	-0.0014*
	(0.0002)	(0.0003)	(0.0007)	(0.0009)
sd(EBIT)	0.0133***	0.0089***		
	(0.0007)	(0.0008)		
Earnings/Asset	0.0409***	0.0486***	0.0467***	0.0511***
	(0.0005)	(0.0006)	(0.0013)	(0.0014)
ln(Asset)	-0.0242***	-0.0214***	-0.0412***	-0.0339***
	(0.0001)	(0.0002)	(0.0010)	(0.0011)
β_{s,r} (deviation)	-0.0097***	-0.0069**		
	(0.0025)	(0.0027)		
Payout ratio	0.0338***	0.0331***	-0.0060***	-0.0054***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Tangible assets/Asset	-0.2495***	-0.2510***		
	(0.0011)	(0.0012)		
Fixed Effects	Sect.-FE & Reg.-FE & Year-FE	Sect.-FE & Reg.-FE & Year-FE	Firm-FE	Firm-FE
Clustering	None	None	Firm	Firm
Weight	None	None	None	None
Sample	At least 6 years	Balanced panel	At least 6 years	Balanced panel
Observations	1,512,449	1,243,475	1,543,338	1,268,913
R2	0.14	0.14	0.79	0.80
Adjusted R2	0.14	0.14	0.73	0.74

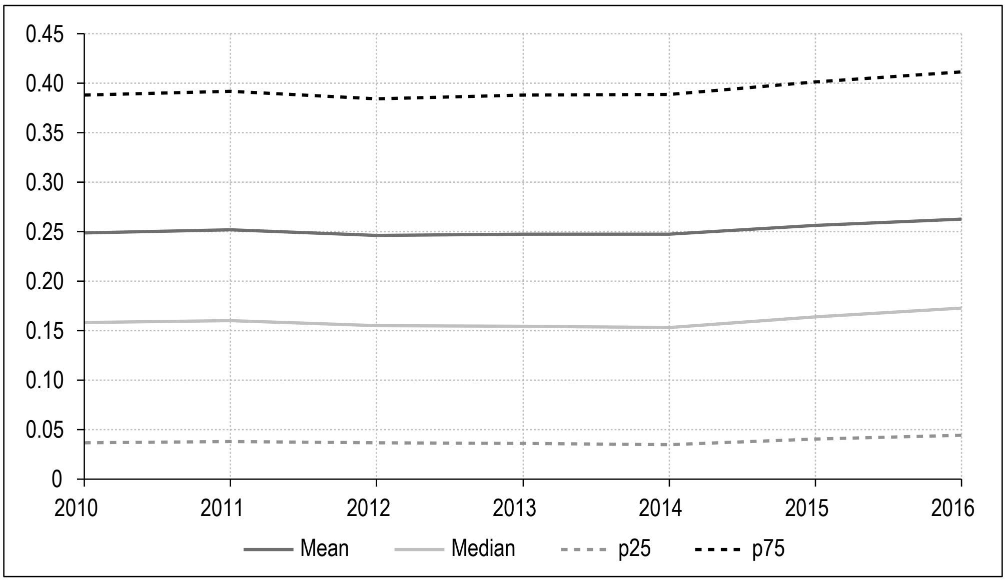
Notes: Variables definitions are given in Table 3. Columns 1 and 2 refer to regression models of Table 6 column 1 with quasi-balanced and balanced sample. Columns 3 and 4 refer to regression models of Table 5 column 1 with quasi-balanced and balanced sample.
Sources: Insee (Esane/LIFI); authors' calculations.

Table A3-1 – **Descriptive statistics**

Variable name	Number of observations	Number of values	mean	sd	q10	q25	median	q75	q90
Cash/Assets	3,665,675	3,111,314	0.20	0.23	0.00	0.03	0.12	0.30	0.55
Cost of Carry (-1)	3,665,675	794,329	3.20	1.77	1.38	2.04	2.75	3.80	5.99
Cost of short-term debt (-1)	3,665,675	454	3.31	1.82	1.38	2.09	2.84	3.95	6.34
Net Working Capital/Asset	3,665,675	3,304,077	0.07	0.75	-0.36	-0.10	0.05	0.26	0.50
Financial debt/Asset(-1)	3,665,675	1,899,255	0.16	0.96	0.00	0.00	0.03	0.20	0.44
Earnings/Asset	3,665,675	2,902,546	-0.02	0.40	-0.19	-0.01	0.02	0.10	0.19
ln(Asset)	3,665,675	1,776,417	5.89	1.75	3.82	4.93	5.90	6.89	7.96
Tangible assets/Asset	3,665,675	2,873,422	0.14	0.20	0.00	0.01	0.06	0.18	0.40
sd(EBIT)	3,665,675	904,318	150	4730	5	13	30	71	172
β_{s,r} (baseline)	3,665,675	285,356	-0.88	0.24	-1.09	-1.02	-0.88	-0.80	-0.69
β_{s,r} (alternative)	3,665,675	285,352	-43.89	44.31	-93.31	-59.48	-35.27	-14.16	0.00
Payout ratio	3,665,675	576,980	6.02	10782	0.00	0.00	0.00	0.00	0.42
Policy Uncertainty	3,665,675	6	250	38	191	224	248	279	310
Hedging needs	3,665,675	531,896	-0.11	0.64	-0.93	-0.68	-0.20	0.43	0.86

Notes: Variables definitions are given in Table 3.
Sources: Insee (Esane/LIFI).

Figure A3.I – **Moments of the financial liquid assets to assets ratios**



Sources: Insee (Esane/LIFI); authors' calculations.

Market Power and Labor Share

Arthur Bauer* and Jocelyn Boussard**

Abstract – Secular trends in market power and labor share have important implications for inequality and allocative efficiency. Studying them requires comprehensive, detailed firm-level data spanning several decades. We leverage a novel database on the universe of French firms between 1984 and 2016 and document a rise in concentration since the early 1990s. Despite a stability of the aggregate labor share, larger firms with lower labor shares gained market shares, especially in industries where concentration increased the most. The markup of the typical firm, considered here as a proxy of its market power, has decreased, but market shares reallocation toward larger firms contributed to an increase in the aggregate markup. In particular, we do not find that the rise in concentration is accompanied by an increase in market power at the top. Finally, we show how taking into account reallocation across firms is essential to understand how the trends in market power have shaped the dynamics of the aggregate labor share in France.

JEL Classification: E10, E23, E25
Keyword: labor share, markups, competition, production function

*Insee and CREST (arthur.bauer@ensae.fr); **European Commission and CREST (boussard.jocelyn@ensae.fr)

Acknowledgements – We thank the economic studies, archives, and firm-level statistics departments at Insee and in particular Kim Anh Fabre, Pascal Lucas, Christelle Le Borgne and Emmanuel Gros for facilitating our access to and understanding of the data. We thank Elizaveta Archanskaia, Sabien Dobbelaere, Dominique Goux, Maxime Gueuder, Francis Kramarz, Edouard Jousselin, Rémy Lecat, Margarita Lopez-Forero, Jacques Mairesse, Rémi Monin, Ludovic Panon, Sébastien Roux, Olivier Simon, and participants at Insee, Banque de France, ECFIN, Collège de France and CREST Seminars for their useful comments. Finally, we thank also two anonymous reviewers.

Received in October 2018, accepted in April 2020.

Citation: Bauer, A. & Boussard, J. (2020). Market Power and Labor Share. *Economie et Statistique / Economics and Statistics*, 520-521, 125–146. https://doi.org/10.24187/ecostat.2020.520d.2034

Large and productive superstar firms have been gaining market shares in many advanced economies, and the rise of their market power, measured either through their markup or their profitability, has been the focus of attention in many recent works. De Loecker *et al.* (2020) have documented an increase in top firms' market power in the US that is large enough to have important macroeconomic consequences. They find that the weighted average markup in the United States rose from 21% above marginal cost at the beginning of the 1980s to around 61% now. Autor *et al.* (2020) also document a rise of the weighted average markup in the US. Gutiérrez & Philippon (2018) argue that European markets are more competitive, and exhibit lower levels of concentration, lower excess profits and lower barriers to entry, which raises the question of whether the secular trends mentionned above are specific to the US. We use detailed firm-level administrative data on the universe of French firms to document facts about market power and labor shares in France.

These questions are important for inequality concerns. One of the important macroeconomic implications of a rise in market power is a decline in the aggregate share of income going to workers. Given that there is ample evidence that labor is more evenly distributed than capital (Garbinti *et al.*, 2018; Piketty *et al.*, 2018) or firm ownership (Bauer *et al.*, 2018), a decline in the aggregate labor share is a possible driver of inequality. Important work has shown that the aggregate labor share has indeed been declining in a wide range of countries (Karabarbounis & Neiman, 2014; Elsby *et al.*, 2013; Grossman *et al.*, 2018). Using aggregate data, Barkai (2020) and Boussard & Lee (2020) show that both the labor and capital shares have declined in the United States and many advanced economies, while measures of the profit share have increased. Looking more closely at firm-level data, Kehrig & Vincent (2018) and Autor *et al.* (2020) show that the labor share of the typical firm has actually increased, while the aggregate fall is attributable to reallocation from high- to low-labor share firms.

Market power trends have also important but ambiguous consequences for allocative efficiency. Baqaee & Fahri (2020) show that a reallocation of market shares to high-markup firms as shown in Autor *et al.* (2020) increases efficiency, but an increase in markup dispersion as shown in De Loecker *et al.* (2020) reduces efficiency. Market power also has important but ambiguous dynamic implications: while lower competition may lead firms to under-invest (Gutiérrez & Philippon, 2017), the relationship between competition and innovation depends on the initial level of competition (Aghion *et al.*, 2005).

Understanding the underlying micro-structural transformations behind these aggregate trends is crucial to identify their possible explanations such as changes in the competitive environment and changes in technology. For instance, Bonfiglioli *et al.* (2019) and Panon (2020) show that national firms compete in markets that are increasingly global, which reduces firm-level markups but benefits larger firms, and Melitz (2003) and Mayer *et al.* (2014) show that international competition causes reallocation toward top producers. Recent work (Autor *et al.*, 2020; Van Rennen, 2018) argues that technological change, such as the growth of platform competition in digital markets, may have caused reallocation from small to large firms that could lead to dominance by a small number of firms. Lashkari *et al.* (2019) find that the rise of Information Technology has disproportionately benefited larger firms.

We use France as a laboratory to study the link between variations in industry concentration and firm-level outcomes, and provide evidence on the sources of market power variations. France is an interesting case because the labor share appears to have been stable or increasing over the past decades, in contrast to the US (see Figure I). We document important facts about secular trends in France that are similar to what has been documented for other advanced economies. When we decompose the labor share variations in France since the 1990s, we find that there was an important reallocation of market shares from firms with high labor shares to firms with low labor shares, which tend to be larger. This reallocation is correlated with a rise of industry concentration, measured through a wide range of proxies. However, labor shares have on average increased at all points of the distribution, a development that has offset the effect of reallocation and explains why the aggregate labor share in France was broadly stable over this period.

To assess the extent to which firm-level market power dynamics have played a role in explaining the diverging trends of firm-level labor share in France and the US, as opposed to other explanations like technological change, we estimate firm-level markups and output elasticities using a flexible production function that allows variations in the marginal product of inputs both across firms and time periods. We follow De Loecker & Warzynski (2012) and first estimate firm-level elasticities of value added to labor and capital, and then recover markups by assuming that firms minimize their costs and that labor is a flexible input. We rely on unique and comprehensive administrative data covering the universe of French firms.

Importantly, we find no evidence that the rise in concentration translated into an increase in firm-level market power. We find that there is substantial heterogeneity in markups, and that markups are increasing with firm size. We also find that much of the increase in firm-level labor shares is attributable to decreases in firm-level markups. All in all, high-markup firms gained market shares while the markup of the typical firm decreased, which indicates both an improvement in allocative efficiency and a decrease in firm pricing power. We show that these two facts about reallocation are strongly correlated with the rise in concentration at the industry level.

Our paper contributes to the macroeconomic literature that documents a number of important secular trends that have recently swept across advanced economies. A number of recent papers have documented growing industry concentration and within-industry dispersion in firm outcomes (Andrews *et al.*, 2016; Berlingieri *et al.*, 2017; Song *et al.*, 2018; Card *et al.*, 2013). In parallel, there is a large body of evidence on a global fall in the labor share across many industries (Elsby *et al.*, 2013; Karabarbounis & Neiman, 2014, 2018; Grossman *et al.*, 2018; Barkai, 2020; Boussard & Lee, 2020). We show that concentration and the market power of top firms are not necessarily correlated, even though at the aggregate level the reallocation of market shares toward high-markup firms contributes to a rise in the aggregate markup. Our findings that *(i)* firm-level markups have decreased and *(ii)* reallocation towards high-markup firms (reflecting a rise in concentration) contributes to a rise in the aggregate markup, are consistent with Autor *et al.* (2020). However, in France, the decrease in firm-level markups is larger, and the reallocation effect does not offset it.[1] This difference is also consistent with evidence in Gutiérrez & Philippon (2018) that European markets have become more competitive than US markets.

The rest of the paper is organized as follows. Section 1 presents our theoretical framework, Section 2 presents our strategy for estimating firm-level markups, Section 3 presents the data that we use to implement this strategy, Section 4 documents important changes in the labor share and concentration in France, and Section 5 presents our results on markups in France.

1. Theoretical Framework

In this section, we provide a general theoretical framework to map variations of the aggregate labor share to variations of firm-level market power, input elasticities and market shares. Consider an industry with N firms indexed by i. Consistently with a wealth of evidence and in the spirit of canonical models (Melitz, 2003; Hopenhayn, 1992), we assume that firms have heterogeneous exogenous productivity Ω_{it} and have access to a common production technology $\mathcal{Q}(.)$ defined as $Y_{it} = \mathcal{Q}(\Omega_{it}, L_{it}, K_{it})$, that they use to produce value added Y_{it}, using variable labor input L_{it}, and capital stock K_{it}. We assume that adjusting the capital stock is subject to cost $\mathcal{C}_a(.)$, which depends only on the current and previous levels of capital, and crucially not on variable inputs levels. The sum of discounted costs of the firm is:

$$\mathcal{V}(\mathrm{Z}_{it}) = \min_{\mathrm{X}_{it}} \mathcal{C}(\mathrm{X}_{it}, \mathrm{Z}_{it}) + \beta \mathbb{E}\left[\mathcal{V}(\mathrm{Z}_{it+1})\right]$$
$$\text{s.t } \mathcal{Q}(\Omega_{it}, \mathrm{X}_{it}) = Y_{it}$$

where $\mathcal{C}(.)$ is the total cost of the firm, $\mathrm{X}_{it} = (L_{it}, K_{it})$ refers to inputs, and Z_{it} to variables that are exogenous to the choice of the firm at time t, such as previous year capital stock, productivity and input prices.

The Lagrangian associated with the right-hand side of the Bellman equation is defined as:

$$\mathcal{L}(\mathrm{X}_{it}, \xi_{it}, Y_{it}, \mathrm{Z}_{it}) = W_{it} L_{it} + r_{it}\left(K_{it} + \mathcal{C}_a(K_{it}, K_{it-1})\right) + F_{it} + \beta \mathbb{E}\left[\mathcal{V}(\mathrm{Z}_{it+1})\right] - \xi_{it}\left(\mathcal{Q}(\Omega_{it}, \mathrm{X}_{it}) - Y_{it}\right)$$

where W_{it} is the wage, r_{it} is the user cost of capital, F_{it} is an exogenous fixed cost, and ξ_{it} is the Lagrange multiplier. The first-order conditions at the optimal choice of inputs (X^*_{it} and ξ^*_{it}) imply that:

$$\nabla L(\mathrm{X}^*_{it}, \xi^*_{it}, Y_{it}, \mathrm{Z}_{it}) = 0 \tag{1}$$

where ∇ denotes the gradient vector of partial derivatives with respect to inputs. Applying equation (1) to the flexible labor input yields the following cost-minimization condition linking the wage and marginal product of labor:

$$\frac{\partial \mathcal{L}}{\partial L}(\mathrm{X}^*_{it}, \xi^*_{it}, Y_{it}, \mathrm{Z}_{it}) = W_{it} - \xi^*_{it} \frac{\partial \mathcal{Q}}{\partial L}(\Omega_{it}, \mathrm{X}^*_{it}) = 0$$

The output elasticity with respect to the labor input L, $\theta_{l,it}$, can therefore be expressed at the optimum as:

$$\theta_{l,it} \equiv \frac{L^*_{it}}{Y_{it}} \frac{\partial \mathcal{Q}}{\partial L}(\Omega_{it}, \mathrm{X}^*_{it}) = \frac{1}{\xi^*_{it}} \frac{W_{it} L^*_{it}}{Y_{it}} \tag{2}$$

1. *Possible interpretations of these difference are that the market power of French firms is more sensitive to the underlying cause, for instance if French firms are more exposed to globalization or to competition on internet platforms than US firms, or if the productivity gap between top French firms and laggards is not as large as for top US firms.*

Using the first order conditions in equation (1) to express the optimal choice of inputs X_{it}^* and ξ_{it}^* as functions of output Y_{it} and exogenous variables Z_{it}, we derive the optimal total cost as a function of output and exogenous variables:

$$\mathcal{C}^*\left(Y_{it}, Z_{it}\right) = \mathcal{C}\left(X_{it}^*\left(Y_{it}, Z_{it}\right), Z_{it}\right)$$

At the optimum, the Lagrangian is equal to total cost, and from the envelop theorem it follows that the marginal cost is equal to the Lagrange multiplier ξ_{it}^*:

$$\frac{\partial \mathcal{C}^*}{\partial Y}\left(Y_{it}, Z_{it}\right) = \frac{\partial \mathcal{L}^*}{\partial Y}\left(Y_{it}, Z_{it}\right) = \frac{\partial \mathcal{L}}{\partial Y}\left(X_{it}^*, \xi_{it}^*, Y_{it}, Z_{it}\right) = \xi_{it}^*$$

Dropping for simplicity the superscript * to denote optimal variables, we define the markup as the ratio of the firm's output price P_{it} to its marginal cost:

$$\mu_{it} = \frac{P_{it}}{\xi_{it}} \quad (3)$$

The markup captures the degree of pricing power of the firm, and is a widely used measure of firm-level market power. As noted by De Loecker & Warzynski (2012), this expression is robust to various static price setting models, and does not depend on any particular form of price competition among firms. The markup itself will, however, depend on the specific nature of competition among firms. Moreover, it follows from equations (2) and (3) that the markup is defined as the elasticity of output with respect to the labor input, divided by the share of this labor costs in total firm revenue, *i.e* the labor share λ_{it}:[2]

$$\mu_{it} = \theta_{l,it} \frac{P_{it} Y_{it}}{W_{it} L_{it}} \equiv \frac{\theta_{l,it}}{\lambda_{it}} \quad (4)$$

In what follows, we map the aggregate labor share into firm level markups, and the output[3] elasticity of labor. First, we define the aggregate labor share Λ_t as the value added weighted average of firm-level labor shares:

$$\Lambda_t \equiv \frac{\sum_i W_{it} L_{it}}{\sum_i P_{it} Y_{it}} = \sum_i S_{it} \lambda_{it} \quad (5)$$

where $S_{it} = \frac{P_{it} Y_{it}}{\sum_i P_{it} Y_{it}}$ is the market share of firm i.

From equation (4) we know that the labor share is the product of the output elasticity of labor and the inverse markup:

$$\lambda_{it} = \theta_{l,it} \mu_{it}^{-1} \quad (6)$$

We decompose the output elasticity of labor $\theta_{l,it}$ into a component stemming from returns to scale, which tells us how much output expands when all inputs increase proportionally, and a component stemming from the labor intensity of the production process relative to capital:

$$\theta_{l,it} = \underbrace{\theta_{l,it} / \left(\theta_{l,it} + \theta_{k,it}\right)}_{\text{Labor intensity}} \times \underbrace{\left(\theta_{l,it} + \theta_{k,it}\right)}_{\text{Returns to scale}} \equiv \alpha_{it} \gamma_{it} \quad (7)$$

noting that when α_{it} is high the production process is intensive in labor relative to capital. It follows from equations (5), (6) and (7) that the aggregate labor share can be expressed as a function of firm level labor intensity, returns to scale, and markups:

$$\Lambda_t = \sum_i S_{it} \alpha_{it} \gamma_{it} \mu_{it}^{-1} \quad (8)$$

We compute the aggregate markup M_t as the value added weighted harmonic average of firm-level markups:

$$M_t \equiv \left[\frac{\sum_i P_{it} Y_{it} \mu_{it}^{-1}}{\sum_i P_{it} Y_{it}}\right]^{-1} = \left[\sum_i S_{it} \mu_{it}^{-1}\right]^{-1}$$

2. Estimation Procedure

In this section, we describe the procedure to recover estimates of firm-level output elasticities of labor and capital; together with firm-level labor and market shares observed in the data, this allows us to compute the contribution of markups, labor intensity, and returns to scale to the aggregate labor share.[4]

To recover markup from production data, we rely on equation (4). This framework is particularly convenient to analyze the evolution of markups in the long run because it does not require observing consumer-level attributes to estimate demand elasticities. Second, it makes no assumption on firms pricing behavior and competition environment. It only requires two assumptions: firms minimize production cost and freely adjust at least one variable input.

We can directly observe firm-specific input shares in production data, but output elasticities are unobserved. Because these elasticities can vary across time and firms, we estimate a flexible production function, with a minimum number of parametric restrictions. In what follows, we assume that firms belonging to a particular industry j share the same technology $f_j(.)$, using labor and capital to generate value added. Moreover, we assume that productivity is Hicks-neutral and evolves according to an AR(1)

2. It is important to note that equation (4) only applies to inputs that are freely adjustable, at least at the margin and that input prices are exogeneous to the firm choices. Section C2 in the Online Appendix discusses the sign of the wedges that arise from relaxing one of these assumptions. Link to the Online Appendices at the end of the article.

3. Actually, the value added. The two terms are used interchangeably hereafter.

4. We abstract from input-output linkages by considering value added production function. Baqaee & Fahri (2020) show that input-output linkages are important for the propagation of productivity shocks, and Grassi (2017) shows that they matter for market power in the case oligopolistic competition.

Markov process. For firm i in industry j, our empirical model is given by:

$$y_{it} = f_j\left(k_{it}, l_{it}\right) + \omega_{it} + \varepsilon_{it} \quad (9)$$

$$\omega_{it} = \rho_{jt}\omega_{it-1} + \eta_j + \nu_j t + \xi_{it} \quad (10)$$

where y_{it} stands for the logarithm of value added of firm i at time t, and l_{it} and k_{it} are the logarithms of employment and capital stock. Productivity ω_{it} is Hicks-neutral, ε_{it} is an i.i.d measurement error, and ξ_{it} is the i.i.d innovation to productivity. Steady-state productivity η_j and time trend ν_j are common across firms in industry j in period t.

One issue that prevents us for simply running Ordinary Least Squares (OLS) on equation (9) is that we do not observe productivity ω_{it} but firms have information about their productivity when they choose their inputs. ω_{it} is therefore correlated with k_{it} and l_{it} and OLS estimates are biased. In what follows, we make the following standard assumptions regarding the timing of firm decisions:

Assumption 1 (Information Set) – The firm's information set at t, i.e. I_t, includes current and past productivity shocks $\left\{\omega_{i\tau}\right\}_{\tau=0}^{t}$ but does not include future productivity shocks $\left\{\omega_{i\tau}\right\}_{\tau=t+1}^{+\infty}$. Measurement errors μ_{it} satisfy $\mathbb{E}\left[\mu_{it} \mid I_t\right] = 0$. The productivity process defined in equation (10) is known to firms and stochastically increasing in ω_{it-1}.

Assumption 2 (Input Choices) – Labor and capital inputs used at time t are chosen with information set I_t.

These assumptions are straightforward: firms do not observe ω_{it} until time t, but the Markov process defines what the firm knows about the distribution of future productivity shocks. To control for unobserved productivity, we rely on an approach usually called dynamic panel estimation (Blundell & Bond, 2000). We use the AR(1) structure of the productivity process to write current value added as:

$$y_{it} = \rho_{jt} y_{it-1} + \left(f_j\left(k_{it}, l_{it}\right) - \rho_{jt} f_j\left(k_{it-1}, l_{it-1}\right)\right) + \eta_j - \nu_j t + u_{it}$$

where the composite error $u_{it} = \xi_{it} + \varepsilon_{it} - \rho\varepsilon_{it-1}$ is zero mean conditional on information set I_{t-1}, by assumptions 1 and 2. Conditioning on a set of instruments included in I_{t-1}, we recover the parameters of the production function and productivity process with a GMM two-step estimation. Our moment conditions can be written as:

$$E\left[u_{it} \mid I_{t-1}\right] = E\Big[y_{it} - \rho_{jt} y_{it-1} - \left(f_j\left(k_{it}, l_{it}\right) - \rho_{jt} f_j\left(k_{it-1}, l_{it-1}\right)\right) - \eta_j - \nu_j t \mid I_{t-1}\Big] = 0 \quad (11)$$

We assume that technology $f_j(.)$ in sector j is a translog production function of capital and labor:

$$f_j\left(k_t, l_t\right) = \beta_{l,jt} l_{it} + \beta_{k,jt} k_{it} + \beta_{ll,jt} l_{it}^2 + \beta_{kk,jt} k_{it}^2 + \beta_{lk,jt} l_{it} k_{it}$$

and we use past values ω_{it-1}, l_{it-1}, m_{it-1}, k_{it-1} and higher order combinations of those terms, a time trend t and a constant as instruments in equation (11). From the estimates of the parameters of the production function, we compute the firm-level output elasticity of labor and capital for firm i in year t as:

$$\theta_{l,it} = \beta_{l,j} + 2\beta_{ll,j} l_{it} + \beta_{lk,j} k_{it}$$

$$\theta_{k,it} = \beta_{k,j} + 2\beta_{kk,j} k_{it} + \beta_{lk,j} l_{it}$$

From equation (7), we retrieve firm-level labor intensity and returns to scale.

Previous studes estimating markups with production data have often estimated production functions on the proxy variable method. This method relies on a non-parametric estimation of unobserved productivity ω_{it} from observed variables using the assumption that some proxy variable, either investment (Olley & Pakes, 1996) or intermediate input demand (Levinsohn *et al.*, 2003; Ackerberg *et al.*, 2015), is an invertible function only of other inputs and productivity. However, this approach is not valid if the proxy variable is also a function of some unobserved shock, such as an input cost shock to all inputs, or a demand shock. Let us define intermediate input demand m_{it} as a function of capital, labor, productivity, and some unobserved shock d_{it}:

$$m_{it} = m\left(\omega_{it}, k_{it}, l_{it}, d_{it}\right)$$

Assuming that this function is invertible in ω_{it} and using equation (9), one can write value added y_{it} as an unknown function of inputs and the unobserved shock:

$$y_{it} = f_j\left(k_{it}, l_{it}\right) + \omega\left(m_{it}, k_{it}, l_{it}, d_{it}\right) + \varepsilon_{1,it}$$
$$= g\left(m_{it}, k_{it}, l_{it}, d_{it}\right) + \varepsilon_{1,it}$$

Ignoring the unobserved shock, and using assumption (1) that ε_{it} is independant from input choices, we can obtain a non parametric estimate $\hat{g}_{it}$ of $g(.)$ that is a high-order polynomial in m_{it}, k_{it}, and l_{it}, but not of d_{it}:

$$y_{it} = \hat{g}_{it} + \hat{\varepsilon}_{it}$$

where the residuals $\hat{\varepsilon}_{it}$ are correlated with d_{it}. In practice, when we apply this procedure, we find that the residuals are not i.i.d. As Doraszelski & Jaumandreu (2019) have recently discussed, d_{it}, as ω_{it}, should also be recognized as potentially correlated with the error term. If so, the instruments used in the second stage of the proxy variable method are not consistent.

3. Data

To carry out our empirical analysis we rely on several sources of micro data produced by the French Institute of Statistics (Insee), covering the universe of French firms spanning the 1984-2016 period. These data are, among other uses, one of the main sources of the elaboration of national accounts. Our sources are gathered out of the universe of firms' tax returns and provide balance sheet, income, and cost information at the firm level, as well as employment, the industry in which the firm operates, the type of legal entity (micro-firms, sole proprietorship entities, or limited liability companies and corporations) and the tax regime to which it is affiliated (micro-regime, simplified regime, or normal regime).

From 1984 to 2007, we rely on the SUSE sources (*Système Unifié de Statistiques d'Entreprises*), gathering information from firms affiliated to two tax regimes, the BRN regime (*Bénéfice Réel Normal*) and RSI regime (*Régime Simplifié d'Imposition*). These files allow to distinguish between payments to labor, material inputs, other intermediary inputs, and investment, and provide information of the book value of capital of the firm and total employment. Hence, they have been widely used in previous research (di Giovanni *et al.*, 2014; Caliendo *et al.*, 2015).

From 2008, we rely on ESANE (*Élaboration des Statistiques Annuelles d'Entreprises*), a dataset that results from the unification of the previous SUSE data and the Annual Surveys of Firms that were conducted each year for broad sectors of industries. Because there is some overlap of information between tax returns and surveys, Insee applies an algorithmic process to reconcile diverging information. To construct our panel of firms we exclude from the post-2008 data firms affiliated to the micro-BIC regime.[5] Moroever, we restrict our analysis to legal units with a unique and valid identifier number.[6]

We focus on market sectors, excluding agriculture because our sample does not cover well firms in that sector.[7] We also exclude real estate and finance, because we focus on the production side of value added distribution among workers and owners of capital and firms. There are 5.7 million firms in our sample, 3.7 million of which have at least one employee. Finally, we rely on industry-level data from KLEMS (Van Ark, 2017) for information on investment and output prices to compute deflated values for value added and capital stocks. Others details on the data are provided in Appendix 1.

3.1. Overview of the Data

Table 1 describes the main variables that we use in our empirical analysis. Our sample of 3.7 million firms with at least one employee spans over 33 years, and contains 27 millions firm-year observations. The average sales are 2.6 million euros, the average number of employees is 14, and the average value of the capital stock is 1.3 million euros. These data are highly skewed: the median level of sales is 285 thousand euros, median number of employees is 3, and median capital stock is 76 thousand euros. This reflects the fact that our data are nearly exhaustive and include many small firms. For firms that report non missing investment, the average reported value is 185 thousand euros, and the median investment is 4 thousand euros, which also partly reflects the fact that investment is lumpy.[8] The average labor share in our sample, computed as the ratio of the sum of the wage bill and payroll taxes to value added, is 75%, close to the median at 74%.[9]

3.2. Aggregate Labor Share

Figure I reports the ratio of compensation of employees, including payroll taxes, to total value added in the macro and micro data, from 1984 to 2016. The aggregate labor share in our sample is lower than the average firm-level labor share. As discussed below in Section 4, larger firms have a lower labor share, which brings down the weighted average labor share. In the sample of firms with at least one employee on which we rely in the rest of the paper, the aggregate labor share decreases from 69.3% in 1984 to 64.7% in 2000, and then increases back to a level close to its initial level, reaching 69.1% in 2016.

5. *An extremely simplified regime introduced in 2008 applicable to very small firms, whose total sales do not exceed 170 thousand euros if the firm operates within the real estate and trade sectors, or 70 thousand euros otherwise. This regime has been widely used by free-lance workers who do not report any capital nor employees.*

6. *A firm is defined as a legal unit with a unique SIREN identifying number. In ESANE, legal units belonging to the same conglomerate are brought together and their accounts are consolidated (Deroyon, 2015). We do not consolidate and keep the underlying legal units as separate firms.*

7. *The market sectors are total economy excluding public administrations, healthcare, and education. The low coverage of agriculture is due to the fact that firms of this sector are mostly affiliated to a tax regime that is not included in the micro-BIC, BRN and RSI regimes.*

8. *The mean of the average firm investment across years is 140 thousand euros and the median of the average firm investment across years is 8 thousand euros.*

9. *Section C1 in the Online Appendix shows that our data is very representative of the market economy, accounting for 87% of total labor costs, 84% of total value added, with little variations over time.*

Table 1 – **Summary statistics**

	Observations	Mean	Median	St.dev
Sales	27,543,090	2,642.6	284.6	77,556.3
Gross output	27,517,472	1,818.5	203.7	69,157.5
Value added	27,517,472	730.0	111.3	32,121.5
Labor costs	27,517,428	507.8	81.0	18,092.5
Labor share	27,334,884	75.1	74.1	33.6
Employment	27,360,292	14.1	3.0	471.6
Intermediary inputs	27,517,477	1,088.5	80.2	46,270.4
Investment	19,814,136	185.1	4.0	19,200.4
Capital book value	27,507,848	1,305.8	76.0	168,003.0

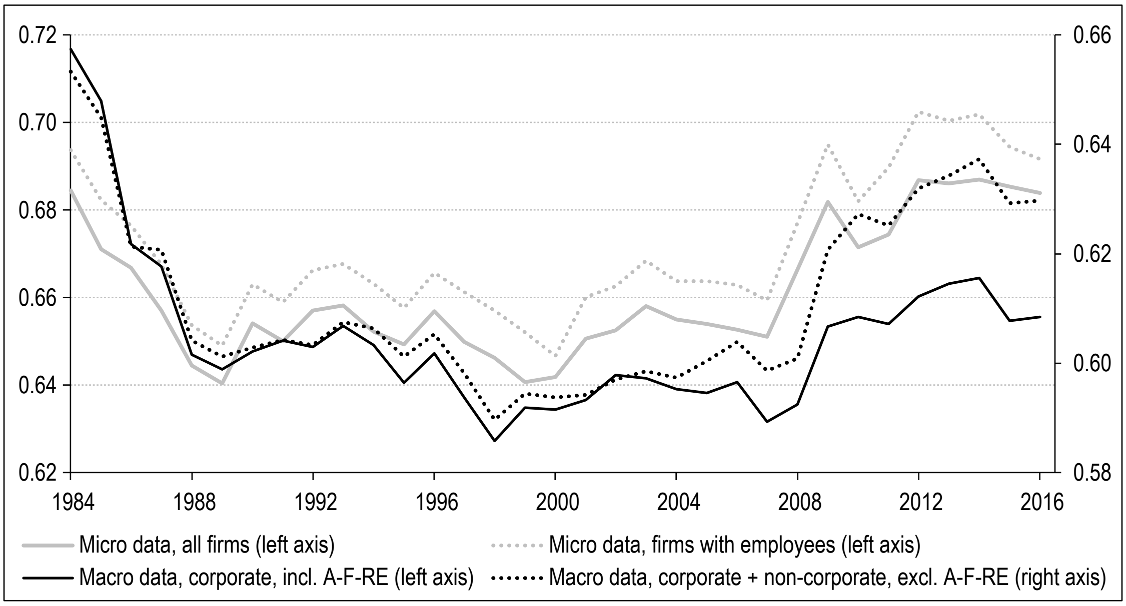
Note: This table presents the main descriptive statistics for the firms in the sample. Values are in thousand euros, except employment which is the number of full-time equivalent salaried workers and the labor share expressed in percentage of value added.
Sources and coverage: Insee, SUSE and ESANE. The sample includes all firms with non zero employment in the corporate market sectors, excluding agriculture, finance and real estate..

The aggregate level is on average 67.1% over the period. Aggregate data in principle also includes firms that have no employee, and doing so in our micro data decreases the aggregate level of the labor share by around 1 percentage point: it stands at 66.1% of value added on average over the period, and has the same U-shaped trajectory. This aggregate pattern differs substantially from the decrease of the labor share in the US, discussed by Autor *et al.* (2020), Kehrig & Vincent (2018), while others have argued that France, as many advanced economies, also experienced a secular decrease in the labor share (see e.g Grossman *et al.*, 2018; Karabarbounis & Neiman, 2014). Because of the U-shaped trajectory of the labor share, both in the micro and macro data, we find that conclusions of a secular decline in France are misguided.

Since our sample excludes agriculture, real estate, and finance, there is no available aggregate data for France for this particular sample; however, the aggregate labor share in our data closely matches the aggregate patterns of the labor share that can be measured for similar spheres of activity, both in levels and in variations.

French national accounts provide detailed operating accounts for spheres that are larger than our data in various dimensions. Figure I reports the labor share of the entire corporate sector, including corporations operating in the agriculture, real estate, and finance. Before 2000, the average level of the labor share in the corporate sector, reported by Insee, is the same as the aggregate labor share in our sample including firms with no employees (65.4%). It starts from

Figure I – **Aggregate labor share in France, 1984-2016**



Note: The figure reports the ratio of employee compensation, including payroll taxes, to total value added in the market sectors in France. See Section 3 for details on the different measures.
Sources and coverage: See Table 1.

a slightly higher level in 1984 (71.6%) than our sample estimate (68.4%) and reaches a slightly lower level in 2000 (63.4% as opposed to 64.1% in our sample). After 2000, however, the corporate labor share rises by 2 percentage points, but the labor share in our sample rises by 4 percentage points.

Figure I also reports the total labor share (corporate and non-corporate) excluding agriculture, real estate, and finance. The non-corporate sector is mainly composed of self-employed workers with few salaried workers. As a result, the total labor share reported by Insee is lower – on average 61% over the period, against 66.1% in our data with all firms. Nevertheless, after 2000, and despite this difference in levels, the rise of the total labor share measured with the same industry composition as our data matches the 4 percentage point increase that we observe in our data. One possible explanation of the divergence between the observed labor share of the corporate sector and that of the market economy excluding agriculture, real estate, and finance, as Cette *et al.* (2019) discuss, is that the growing share of the real estate sector, which has a labor share close to zero in total value added contributes negatively to the aggregate labor share of the corporate sector, especially during the housing boom years after 2000.

4. Labor Share and Concentration

In this section, we revisit some important facts about concentration and labor shares in the French context. In particular, we find that the rise in concentration in France is associated with an increase in firm-level labor shares, and a reallocation of market shares towards large and low-labor-share firms.

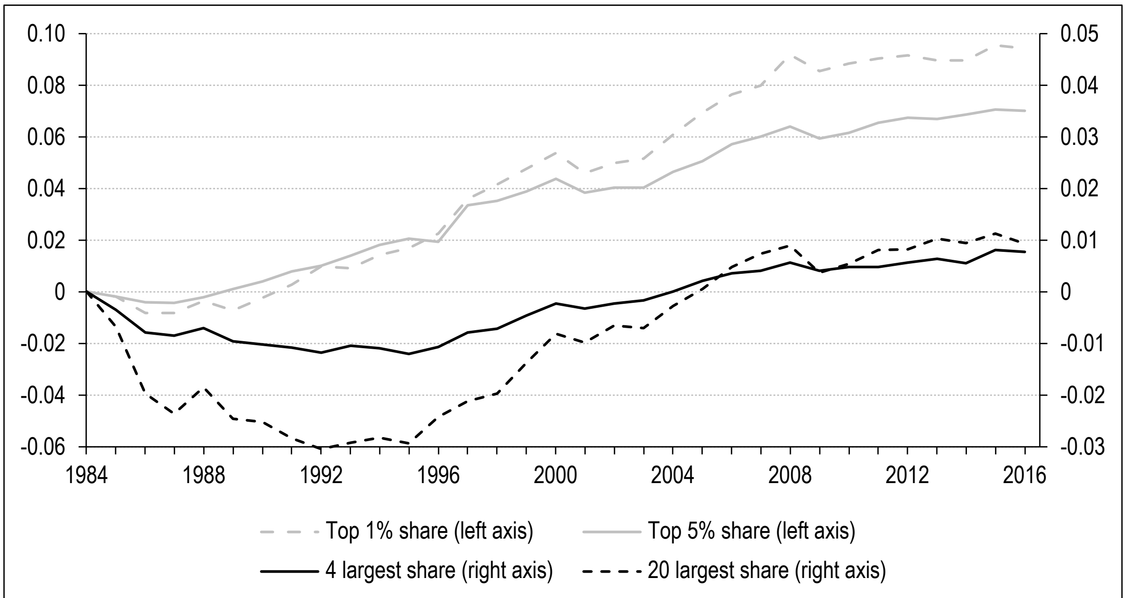
4.1. Rise in Concentration

Figure II reports the cumulative change since 1984 in sales weighted average levels of industry concentration indexes, where each index measures concentration of sales at the 3-digit national industry level. The share of sales of the 1% or 5% largest firms in each industry increased sharply on average since 1984, by 9 and 7 percentage points respectively. The concentration ratios, defined as shares of the 4 and 20 largest firms in each industry, followed a different pattern before 1995 but have increased by close to 4 percentage points each on average since 1995.[10]

Overall, we find that concentration ratios and top shares have increased in more than half of the 211 industries since 1995: the median increase of both concentration ratios is 2 percentage points, and the median increases of the top 1% and 5% shares are 4 and 5 percentage points

10. The median 3-digit industry has around 900 firms in a given year, but because 25% of the industries have more than 5,000 firms, and 25% have less than 200 firm, the number of firms in the top 1% and 5% differs greatly from one industry to the next. The median size of the 3-digit manufacturing industry is around 500 and the median size of the 3-digit non-manufacturing industry is 3,600.

Figure II – **Cumulative change in sales concentration**



Note: The cumulative change in sales concentration is measured across 3-digit industries. Industry changes in concentration are weighted by the share of each industry in total sales.
Sources and coverage: See Table 1.

respectively.[11] These results are consistent with evidence across the US and other OECD countries (CEA, 2016; Autor *et al.*, 2020; Andrews *et al.*, 2016).

4.2. Reallocation of Labor Shares

We build on Kehrig & Vincent (2018) and decompose the variations of the aggregate labor share to understand whether they are driven by variations at the firm level or by composition effects. Figure III reports, for each decile of labor share, the value-added-weighted average labor share and the share of industry value added of firms in that decile, in the first and last five years of the sample. Firms in the lowest decile of labor share accounted for 12% of their industry value added before 1990, compared to 16% in after 2010. The rise in industry shares is verified for four out of the five lowest deciles of labor share, while all five highest deciles of labor share accounted for less of industry value added in 2011-2016 than in 1984-1989. The lines illustrate how the raw distribution of labor shares has shifted upwards: the average labor share of each decile is higher in after 2010 than before 1990. The vertical bars illustrate how low labor share firms gained market share in the last 30 years.

To quantify how these dynamics affect the aggregate labor share, we compute the contributions of industry reallocation, firm reallocation, and firm labor shares to the variations of the aggregate labor share.[12] Figure IV reports the results of this decomposition. Reallocation across industries plays only a minor role in aggregate labor share variations. However, reallocation towards low-labor-share firms contributed to an accumulated 5 percentage points decrease of the aggregate labor share since 1984. This was offset by the upward shift in the labor share distribution, that contributed to a rise of the aggregate labor share of 5 percentage points.

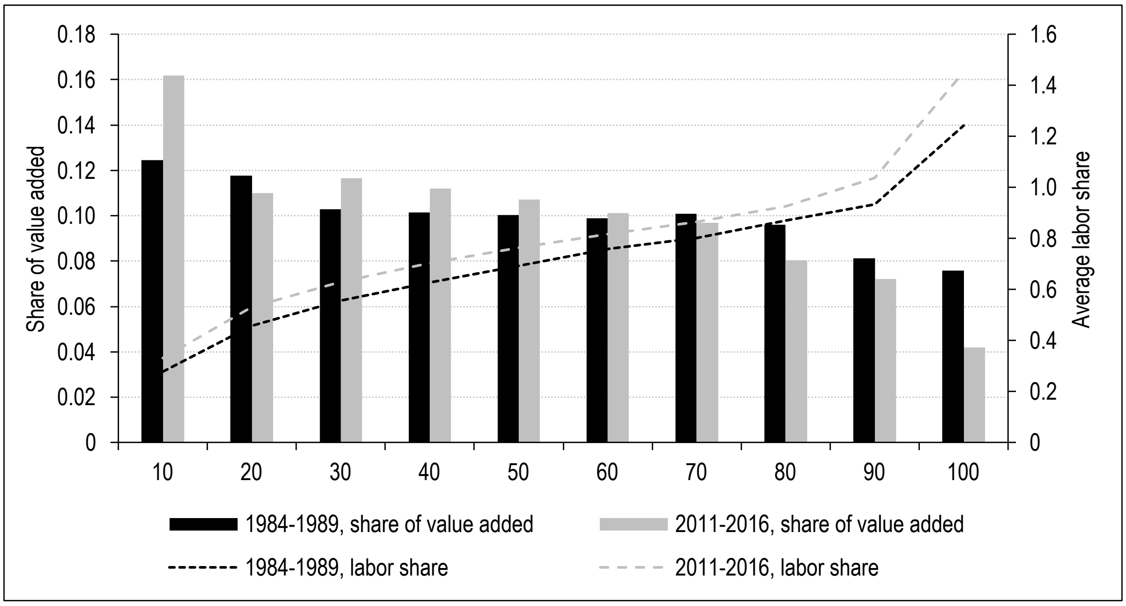
As emphasized by Kehrig & Vincent (2018), this decomposition groups firms into labor shares quantiles, which allows us to compare two static equilibria. It is conceptually distinct from standard within and between firm decompositions, because it abstracts from the contributions of firms' entry and exit. We focus on long term shifts in the joint distribution of labor and value added shares, not on the role of entry nor on the trajectories of specific firms (Section C3 in the Online Appendices discusses firm-level trends).

4.3. Correlation of Rise in Concentration and Reallocation of Labor Shares

We now show that variations in industry concentration are related to these labor share trends. We estimate the industry-level relationship between

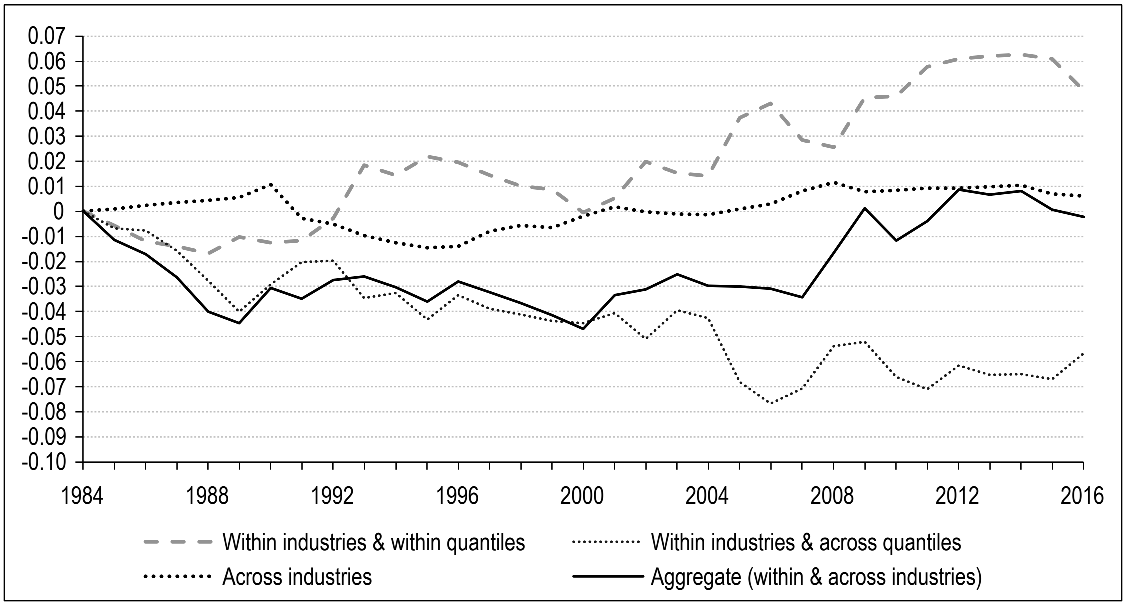
11. Section C5 in the Online Appendices discusses the results in manufacturing and non-manufacturing.
12. The details of decomposition are presented in Appendix 2.

Figure III – **Distributions of labor shares and value added**



Note: Dashed lines correspond to the raw cross-firm distribution of labor shares (scale on the right-hand axis). Vertical bars reflect the share of industry value added of firms in each unweighted decile of labor share (scale on the left-hand axis). To account for industry-specific differences in the joint distributions of labor share and value added, these distributions are averaged across 3-digit industries using value added weights in a given year, and averaged across 5 year periods.
Sources and coverage: See Table 1.

Figure IV – **Decomposition of the aggregate labor share**



Note: The decomposition of the aggregate labor share is described in Appendix 2. Quantiles of labor share are calculated each year within 3-digit industries.
Sources and coverage: See Table 1.

changes in concentration and changes in labor share. We run the following regression:

$$\Delta\lambda_{jt} = \psi_{\lambda}\Delta Conc_{jt} + FE_t + \varepsilon_{jt} \qquad (12)$$

where $\Delta Conc_{jt}$ is the 10-year change of sector j concentration level, proxied by the top 1% and top 5% share of sales, FE_t is a set of time fixed-effects that control for year-specific shocks, and $\Delta\lambda_{jt}$ is the 10 year change in industry j labor share.

Table 2 reports the results. The first two columns show that variation of industry concentration are negatively correlated with variation of industry labor shares. This relationship is significant and holds for all proxies of concentration. We find that a 10 percentage point rise in concentration is associated with a 0.7 to 1.1 decline in the weighted average labor share of the industry. These results are similar to those documented in the US (Autor *et al.*, 2020).

We then consider two components of the 10-year change of the labor share: the cross-quantile contribution to the labor share variation discussed in the previous paragraph, and the evolution of the average labor share of the 5% firms with the lowest labor share within each industry. We

Table 2 – **Correlations between variations in industry-level concentration and labor shares**

	Industry labor share		Across labor share quantiles		Within low labor share quantiles	
Top 1% share	-0.0777		-0.0457		0.0097	
	(0.0123)		(0.0112)		(0.0099)	
Top 5% share		-0.1102		-0.1288		0.0092
		(0.0167)		(0.0150)		(0.0135)
Observations	4,666	4,673	4,665	4,660	4,661	4,664
R2	0.0341	0.0347	0.0290	0.0405	0.0281	0.0292
4 largest shares	-0.0728		-0.0602		0.0772	
	(0.0147)		(0.0133)		(0.0119)	
20 largest shares		-0.1113		-0.1196		0.0615
		(0.0168)		(0.0152)		(0.0137)
Observations	4,649	4,648	4,645	4,645	4,651	4,650
R2	0.0320	0.0388	0.0325	0.0401	0.0366	0.0340

Note: Each estimate is the result of OLS estimation at the 3-digit industry with year fixed-effects. The dependent variable in columns "Industry labor share" is the long-term change of the industry aggregate labor share, defined as the ratio of the sum of firm level compensation and taxes paid on labor over the sum of firm level value added in that industry. The dependant variable in columns "Across labor share quantiles" and "Within low labor share quantiles" are the corresponding contributions to the industry aggregate labor share according to the decomposition described in Appendix 2, where low quantiles are the bottom 5%. The independent variables are the changes in the share of sales for the top 1%, top 5%, 4 largest and 20 largest firms.
Sources and coverage: See Table 1.

use these components as dependent variables in equation (12).

We find that larger increases in concentration are associated with a more negative contribution of value added share reallocation to the aggregate labor share. All coefficients are negative and significant. We also find a positive correlation between change in concentration and change in the average labor share of low labor share firms, defined as firms with a labor share in the bottom 5% of their 3-digit industry. These firms are sometimes referred to in the literature as 'hyper-productive' (Kehrig & Vincent, 2018) or 'superstar' firms (Autor *et al.*, 2020). As we will show next, firms with low labor shares also tend to be larger in our sample. These results suggest that the negative correlation between labor share and concentration is not driven by a decrease in the labor share of 'superstar' firms as they gain market shares.

4.4 Labor Share and Firms' Size

In fact, we show that the negative correlation between reallocation towards low labor share firms and concentration is largely driven by a monotically decreasing relationship (on average) between labor share and firm size. We run the following regression:

$$\lambda_{it} = FE_{size_{it}} + FE_{jt} + \varepsilon_{it} \quad (13)$$

where $FE_{size_{it}}$ is a set of dummies indicating the size class of firm i in industry j in terms of employment at time t, FE_{jt} is a set of interacted fixed effects at the 3-digit industry j and year level.

Figure V presents the results of this regression, considering labor share in value added and in gross output. Relative to 10-20 employee firms, larger firms tend to report lower labor shares even after controlling for industry and year fixed effects. This decreasing relationship is monotonic, at all levels of employment. Labor shares of firms with 50 to 100 employees tend to be 2 percentage points lower than labor shares of 10 to 20 employees firms of the same industry at the same year. For firms with 2,500 to 5,000 employees the gap rises to 5 percentage points considering labor share in value added and to 7 percentage points considering labor share in gross ouput.

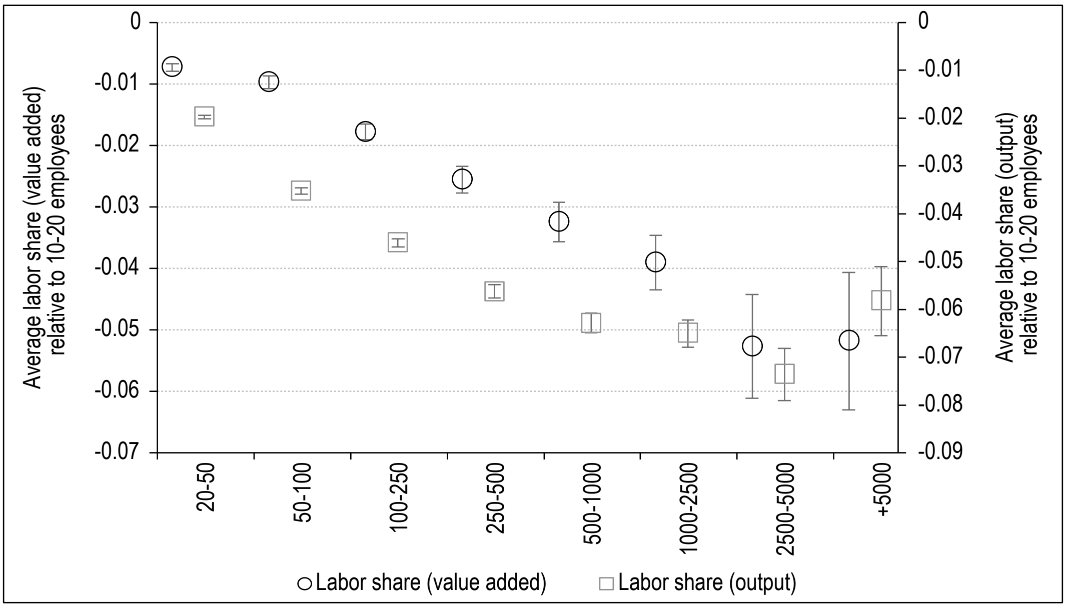
5. Estimation Results

In this section, we first present the results of our estimation procedure, and then show how aggregate and firm-level markups have evolved in France. We document additional facts about market power and concentration, and how variations in market power have contributed to the aggregate labor share, compared to other technological factors.

5.1. Production Function

Table 3 reports the results of rolling estimation of the production function, for the 27 sectors

Figure V – **Labor share by firm size**



Note: The figure reports the conditional average labor share by firm size, with 99% confidence interval. Averages are conditional on a set of flexible fixed effects constructed from the interaction of 3-digit industry codes and year.
Sources and coverage: See Table 1.

Table 3 – **Average output elasticities, rolling estimation**

	θ_l	θ_k	Observations		θ_l	θ_k	Observations
Mining	0.611	0.289	45,698	Gas and electricity	0.697	0.236	22,243
	(0.199)	(0.162)			(0.190)	(0.174)	
Food products	0.754	0.127	1,277,913	Water supply and waste	0.630	0.204	118,249
	(0.052)	(0.104)			(0.178)	(0.146)	
Textiles	0.553	0.135	282,598	Construction	0.611	0.078	4,969,117
	(0.221)	(0.157)			(0.175)	(0.087)	
Wood, paper and printing	0.794	0.044	552,510	Wholesale and retail trade	0.762	0.093	8,502,337
	(0.110)	(0.104)			(0.175)	(0.145)	
Coke and refined petroleum	0.533	0.251	2,472	Transportation	0.840	0.045	988,348
	(0.391)	(0.258)			(0.156)	(0.148)	
Chemicals	0.806	0.163	62,567	Accomodation and food	0.592	0.181	3,076,031
	(0.143)	(0.122)		services	(0.174)	(0.133)	
Pharmaceuticals	0.898	0.072	11,657	Publishing and motion	1.077	-0.001	309,540
	(0.359)	(0.286)		pictures	(0.245)	(0.215)	
Rubber and plastic products	0.763	0.125	245,896	Telecommunications	1.048	-0.035	25,191
	(0.159)	(0.176)			(0.242)	(0.217)	
Basic Metals	0.719	0.111	545,742	ICT	0.921	0.002	324,622
	(0.128)	(0.095)			(0.140)	(0.140)	
Computers and electronics	0.747	0.095	110,072	Legal, accounting	0.843	-0.020	1,499,590
	(0.084)	(0.068)		and engineering	(0.164)	(0.150)	
Electrical equipments	0.766	0.127	50,476	Scientific research	0.856	0.015	30,461
	(0.136)	(0.101)			(0.259)	(0.230)	
Machinery and equipments	0.808	0.094	161,603	Advertising and market	0.867	-0.067	406,636
	(0.137)	(0.069)		research	(0.269)	(0.140)	
Transport equipments	0.834	0.121	71,000	Administrative	0.757	0.039	1,401,753
	(0.180)	(0.156)		and support services	(0.126)	(0.165)	
Other manufacturing products	0.745	0.042	650,254	Total	0.724	0.086	25,744,576
	(0.129)	(0.080)			(0.193)	(0.143)	

Note: Columns θ_l and θ_k report the average estimated output elasticity with respect to each factor of production for the translog production function for all firms. Standard deviations across firms (not standard errors) of the output elasticities are reported in brackets.
Sources and coverage: See Table 1.

of our data. These estimates are obtained by first estimating the parameters of the production function $\beta_j \in \left\{\beta_{l,j};\beta_{k,j};\beta_{ll,j};\beta_{kk,j};\beta_{lk,j}\right\}$ in industry j on 11-year rolling window samples, and then averaging for each firm each year the various estimated output elasticities based on samples that include that year:[13]

$$\beta_{jt}^{rolling} = \frac{1}{11}\sum_{n=-5}^{5}\beta_j^{t+n}$$

where β_j^t is the estimated parameter on the sample restricted to years t-5 to t+5. For the first and last five years of our sample, the average is calculated on fewer estimates. Output elasticities also vary across firms in the same sector. We report, for the different sectors, the average and standard deviation of the elasticities.[14] Because the returns to scale vary across firms, it is possible for many firms in a sector to have increasing returns to scale, while the estimate of the industry average returns to scale is close to 1. On average, the output elasticity of labor in our data is 0.72.

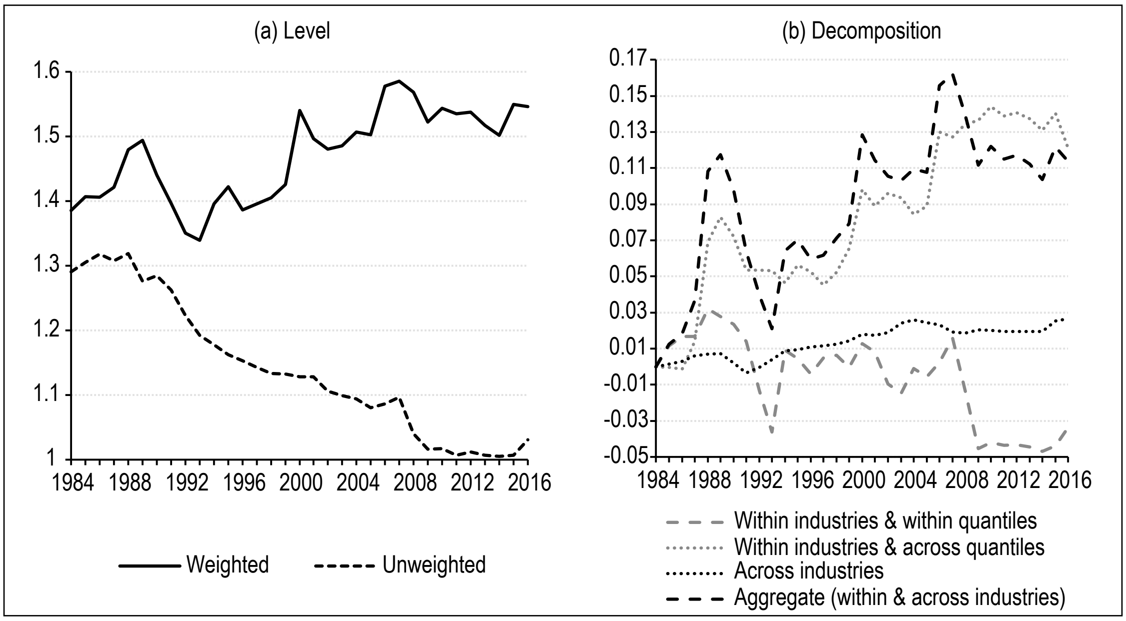
5.2. Aggregate Markup

The left panel of Figure VI reports the variations of the value added weighted and unweighted average markups across all firms in our sample. The unweighted average markup is smaller than the weighted average markup, because firms with larger value added have on average higher markup. We find that the unweighted average markup has decreased in France from 1.3 in 1984 to 1.0 in 2016. The value-added-weighted markup has increased from 1.4 to 1.6.[15]

The right panel of Figure VI shows the decomposition of the aggregate (weighted average) markup into within markup-quantile and

13. We estimate the production for each of the 27 sectors. Each sector includes several 3-digit industries. Section C5 in the Online Appendices reports the results of the non-rolling estimation.
14. We note that a few sectors appear to have negative average capital elasticities or low returns to scale. Section C5 in the Online Appendices reports median output elasticities which are less influenced by outliers.
15. Section C5 in the Online Appendices discusses the results in manufacturing and non-manufacturing. Section C4 of the Online Appendices discusses the results with other estimation methods (non rolling and following the proxy method of Ackerberg et al.*, 2015).*

Figure VI – **Aggregate markup**



Note: The levels of the weighted and unweighted mean markup are based on rolling estimation of a translog value added production function. The decomposition of the aggregate markup is described in Appendix 2. Quantiles of markup are calculated each year within 3-digit industries.
Sources and coverage: See Table 1.

across markup-quantile components. It shows the importance of controlling for industry and disentangling the respective contributions of variations in value added shares holding markup constant or in markup holding value added shares constant to interpret aggregate variation.

The decomposition of the aggregate markup mirrors the decomposition of the aggregate labor share and shows how the within markup-quantile component contributed negatively to the evolution of the aggregate markup, while the cross-quantile component contributed positively. The contribution of reallocation across industries is negligible. Firms with relatively higher markups within narrowly defined industries have been gaining market shares, while the typical firm markup has slightly decreased.

5.3. Markup and Concentration

As for the labor share, we examine whether the observed rise in concentration is correlated with markup variations, on aggregate or along the distribution of markups. To that end we estimate the industry-level relationship between long term changes in concentration and the industry aggregate markup, or the contributions to the aggregate variation. We run the following regression:

$$\Delta\mu_{jt} = \psi_{\mu}\Delta Conc_{jt} + FE_t + \varepsilon_{jt} \qquad (14)$$

where $\Delta\mu_{jt}$ is the 10-year change of sector j aggregate markup level, or one of its contributions according to the decomposition described in Appendix 2.

Table 4 reports the results of the estimation of equation (14). The first two columns show that there is a positive and significant long-term relationship between the evolution of the aggregate markup and the evolution of concentration at the 3-digit industry level. This relationship is significant and holds for all proxies of concentration.

Next, as for the labor share, we ask whether this result is driven by a correlation between the rise in concentration and the shift in value added shares from low to high markup firms. The coefficients of the third and fourth columns of Table 4 are the results of regressions described in equation (14) where the dependent variable is the cross-quantile component to the evolution of aggregate markup, while in the last two columns the dependent variable is the within-quantile component of firms high markups, defined as firms with a markup in the top 5% of their 3-digit industry.They show a positive correlation between the rise in concentration and the cross-quantile component of the evolution of the aggregate markup. As for the labor share, this means that the cross-quantile component contributed more to the rise in markup in those industries that have become more concentrated at the top.

The fifth and sixth columns of Table 4 show no evidence that a rise in concentration is correlated

Table 4 – **Correlations between variations in industry-level concentration and markup**

	Industry markup		Across markup quantiles		Within high markup quantiles	
Top 1% share	0.2640		0.0790		0.0092	
	(0.0257)		(0.0245)		(0.0145)	
Top 5% share		0.3577		0.1460		0.0400
		(0.0353)		(0.0337)		(0.0199)
Observations	4,660	4,660	4,654	4,654	4,663	4,663
R2	0.0569	0.0586	0.0120	0.0140	0.0168	0.0177
4 largest shares	0.2098		0.0995		-0.0536	
	(0.0321)		(0.0298)		(0.0175)	
20 largest shares		0.1702		0.1101		-0.0242
		(0.0372)		(0.0346)		(0.0202)
Observations	4,647	4,646	4,644	4,644	4,650	4,650
R2	0.0482	0.0447	0.0108	0.0112	0.0172	0.0173

Note: Each estimate is the result of OLS estimation at the 3-digit industry with year fixed-effects. The dependent variable in columns "Industry markup" is the long-term change of the industry aggregate markup. The dependant variable in columns "Across markup quantiles" and "Within high markup quantiles" are the corresponding contributions to the industry aggregate markup according to the decomposition described in Appendix 2, where high quantiles are the top 5%. Markups are computed using rolling estimation of a translog production function. The independent variables are the changes of the share of sales of the top 1%, top 5 %, largest 4 and largest 20 firms.
Sources and coverage: See Table 1.

with increases in top markups. The correlations with variations in the top 1% and 5% shares of sales are not significantly positive, the correlations with variations in the shares of the 4 and 20 largest firms are all negative, and significant at the 5% level when concentration is measured with the share of the 4 largest firms.[16] The fact that top markups are not linked with rises in concentration is consistent with theories according to which high productivity firms with higher markups benefit from positive shocks, such as export demand shocks, more than laggard firms, and expand without increasing their markup (see e.g Aghion *et al.*, 2019). However, it is in contrast with results in the US documented by De Loecker *et al.* (2020) where top markups contributed to a third of the overall increase in weighted average markups. Nevertheless, De Loecker *et al.* (2020) do not provide evidence that the rise in top firms' markups is correlated at the industry level with the reallocation component, or with concentration.

5.4. Markup and Size

As for the labor share, we investigate whether markups are increasing with firm size to understand the correlation between the growing share of the largest firms in each industry's total sales and the reallocation of market shares towards high markup firms. To that end, we run the following regression:

$$\mu_{it} = FE_{size_{it}} + FE_{jt} + \varepsilon_{it} \quad (15)$$

where $FE_{size_{it}}$ is a set of dummies indicating in the size class of firm i in industry j in terms of employment at time t, FE_{jt} is a set of interacted fixed effects at the 3-digit industry j and year level.

Figure VII reports the results of this regression. We find that larger firms have higher estimated markups. Firms with more than 5,000 employees have, on average, markups larger by 30 percentage points than firms with 10 to 20 employees within the same 3-digit industry on the same year. This increasing relationship is well observed at all levels of employment, and both for markups obtained with the non-rolling and rolling estimations.

The markup is defined in equation (4) as the ratio of the output elasticity of labor to the labor share. It is important to note that because the output elasticity of labor vary across firms, the markup is not perfectly correlated with the labor share, and therefore the positive relationship between a firm's markup and its size does not flow directly from the negative relationship between its labor share and its size documented in Section 4.4.

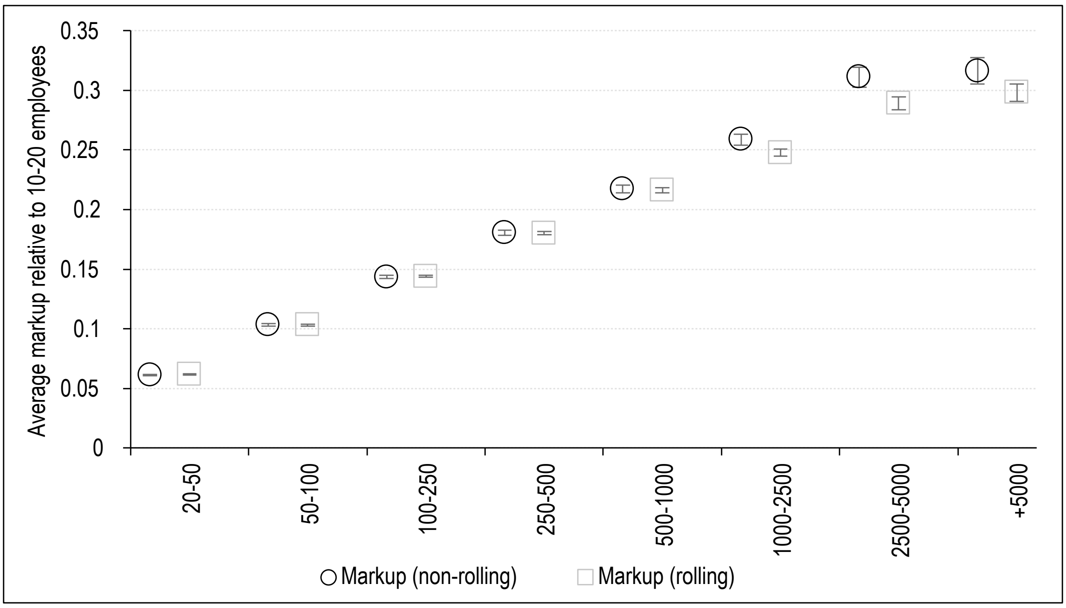
5.5. Link Between Labor Shares and Markups

In this section, we return to the labor share and ask whether variations in firm-level labor share are mainly driven by markups – i.e. are labor shares increasing because markups are decreasing? – or by technology – i.e. are labor shares increasing because production has become more labor intensive?

First, we find that there is a clear negative relationship between firm-level labor shares and markups in France. We run the following regressions:

16. *See Section C5 of the Online Appendix for results limited to manufacturing or non-manufacturing industries.*

Figure VII – **Markup and size**



Note: The figure reports the conditional average markup by firm size, with 99% confidence interval. Averages are conditional on a set of flexible fixed effects constructed from the interaction of 3-digit industry codes and year.
Sources and coverage: See Table 1.

$$\lambda_{it} = \phi\mu_{it} + FE_{ijt} + \varepsilon_{it} \quad (16)$$

where μ_{it} is the markup of firm i in year t, λ_{it} is the labor share, and FE_{ijt} is a set of fixed effect, either industry or firm-level, and year.

Table 5 presents the results of these regressions, and shows that firms with high markup have low labor shares both across industries and across firms within the same industry. We also find that as the markup of the firm grows, its labor share decreases. The absolute value of coefficient ϕ is around 0.3 to 0.5 depending on the estimation: as the markup of the firm increases 10 percentage points, its labor share decreases by 3 percentage points. Finally, as the coefficient of determination of the regression without fixed effects shows, the heterogeneity of markups explains 45% of the heterogeneity of labor shares across firms. The different panels of the table show that this relationship holds statistically and quantitatively for various groups of size.

To extrapolate these firm-level results to the aggregate economy, we need to keep in mind that there is no such a thing as a representative firm in this context. Recall that equations (6) and (7) show that at the level of the individual firm, the labor share is the product of labor intensity, returns to scale and the inverse markup ($\lambda_{it} = \alpha_{it}\gamma_{it}\mu_{it}^{-1}$) but this result does not hold at the aggregate level. From equation (8), we now decompose variations of the aggregate labor share into contributions from labor intensity, returns to scale, and markups, either by taking the "representative firm" approach and computing the contributions of the weighted averages of each component of the aggregate labor share, therefore ignoring the reallocation between firms or, alternatively, by isolating the contribution of reallocation and computing the contributions of the unweighted averages of each component.[17]

The left panel of Figure VIII presents the results of the decomposition for the representative firm. The total variation of the aggregate labor share from 1984 to 2016 is small and positive, and ignoring the role of reallocation, the aggregate markup has contributed negatively to the aggregate labor share, which is consistent with previous evidence that the aggregate markup has increased from 1984 to 2016. The sum of the contributions of labor intensity and returns to scale, in other words the contribution of the weighted average output elasticity of labor, is positive, which would suggest that the French economy has become more 'labor intensive' over the period.

However, taking into account reallocation provides a different picture of underlying determinants of the dynamics of the aggregate labor share in France. The right panel of Figure VIII presents the results of the decomposition isolating the contribution of reallocation. The contribution of reallocation is negative and very large, as we have already showed in Figures IV

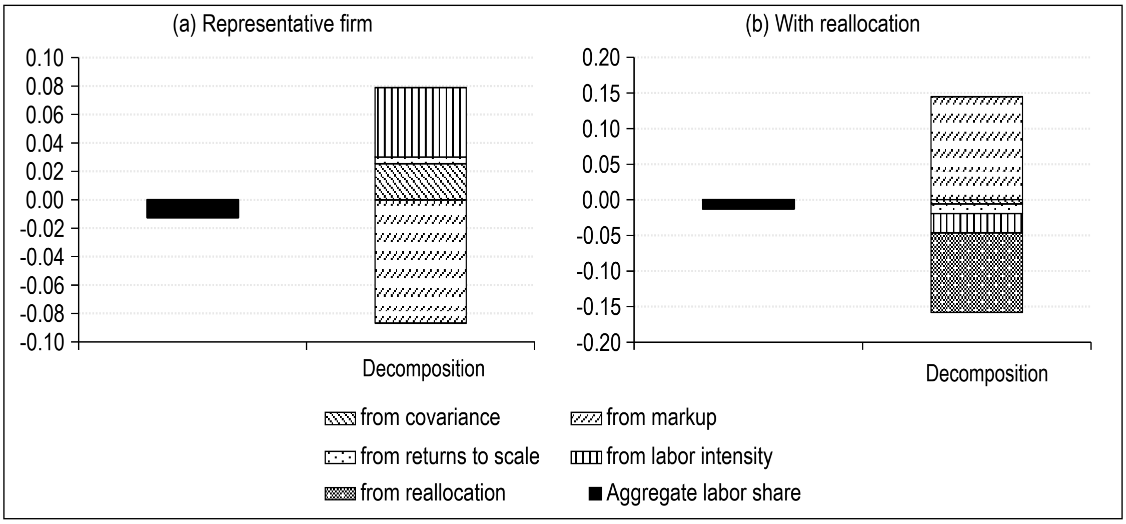
17. See Appendix 3 for details on the decomposition.

Table 5 – **Correlation between labor share and markup**

	Dependent variable: labor share					
	No size threshold			More than 50 employees		
	No FE	Industry FE	Firm FE	No FE	Industry FE	Firm FE
Markup	-0.3173	-0.3520	-0.3370	-0.4070	-0.4351	-0.4797
	(0.0041)	(0.0022)	(0.0027)	(0.0054)	(0.0035)	(0.0044)
Observations	25,554,561	25,554,533	25,092,587	808,003	807,805	789,488
R2	0.407	0.489	0.761	0.493	0.582	0.805
	More than 100 employees			More than 1,000 employees		
	No FE	Industry FE	Firm FE	No FE	Industry FE	Firm FE
Markup	-0.3842	-0.4163	-0.4554	-0.3270	-0.3709	-0.3912
	(0.0053)	(0.0041)	(0.0053)	(0.0050)	(0.0077)	(0.0125)
Observations	398,301	398,018	390,768	26,684	25,305	24,839
R2	0.483	0.594	0.814	0.471	0.710	0.892

Note: Each estimate is the result of OLS estimation of firm level labor share on markups, for four samples: all firms, firms with more than 50 employees, 100 employees, and 1000 employees. Markups are computed using rolling estimation of a translog production function. All columns include year fixed effects. Standard errors are clustered at the 3-digit x year industry level. FE stands for fixed effects.
Sources and coverage: See Table 1.

Figure VIII – **Contributions to the evolution of the aggregate labor share, 1984-2016**



Note: The decomposition of the variation of the agregate labor share is based on translog non-rolling and rolling value added estimation of the production function. See Appendix 3 for details.
Sources and coverage: See Table 1.

and III. Firm-level markups have contributed positively to the aggregate labor share, while firm-level returns to scale and labor intensity had a negative contribution.

* *
*

In this paper, we find no evidence of a rise in firms' market power in France: firm-level markups decreased on average, and the rise in concentration is not correlated with increases in top markups. These facts are however correlated with an important reallocation of market shares towards low-labor share and high-markup firms, which contributed to a rise in the aggregate markup. Because those firms tend to be larger, this reallocation translates into a rise in concentration.

This reallocation of market shares towards large firms is consistent with a wealth of evidence about the increasing differences between firms (Decker *et al.*, 2016a, 2016b, 2016c; Andrews *et al.*, 2016; Karahan *et al.*, 2019). However, the simultaneous rise in concentration and the relative stability of top firm-level markups raises questions about the interpretation of concentration that go beyond the French case. One possible way to explain both the reallocation of market shares towards large firms and the within-firm increase in the labor share would be an increase in winner-take-most competition, as discussed by Autor *et al.* (2020): as consumers become more sensitive to firms' prices, more productive and bigger firms gain market shares but a given firm's market power decreases. The source of this increase in competition could be international competition (Bonfiglioli *et al.*, 2019; Panon, 2020). Since our results hold across broad sectors of the French economy, including non-manufacturing sectors, other factors than international competition could be at play. Technological factors, such as the rise of internet platforms and price comparison websites, may for instance explain why firm-level market power has decreased.

Many predictions of the textbook explanation of a rise in competition are consistent with the evidence provided here. We do not take a stance on the source of market power, and in particular on why there is an increasing relationship between a firm's size and its markup: the price elasticity of demand may decrease with quantity, or large firms may be large enough to influence the equilibrium price, and therefore act strategically. However, in both cases, an increase in competition will have offsetting effects on the markup of large firms: holding size constant, it will tend to decrease their markup, but because of reallocation, these firms will grow and their markup will increase. Qualitatively, it is thus possible to observe a rise in top firms' markups, as De Loecker *et al.* (2020) find for the US, or a stability or decrease, as we find for France. ☐

Link to the Online Appendices: https://insee.fr/en/statistiques/fichier/4997869/ES-520-521_Bauer-Boussard_Online_Appendices.pdf

BIBLIOGRAPHY

Ackerberg, D., Caves, K. & Frazer, G. (2015). Identification Properties of Recent Production Function Estimators. *Econometrica*, 83(6), 2411–2451. https://doi.org/10.3982/ECTA13408.

Aghion, P., Bloom, N., Blundell, R., Griffith, R. & and Howitt, P. (2005). Competition and innovation: an inverted-U relationship. *The Quarterly Journal of Economics*, 120(2), 701–728. https://doi.org/10.1093/qje/120.2.701

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J. & Li, H. (2019). A Theory of Falling Growth and Rising Rents. Federal Reserve Bank of San Francisco, *Working Paper* N°2019-11 https://doi.org/10.24148/wp2019-11

Andrews, D., Criscuolo, C. & Gal, P. (2016). The Best versus the Rest: The Global Productivity Slowdown, Technology Divergence across Firms and the Role of Public Policy. *OECD Productivity Working Paper* N°5. https://doi.org/10.1787/63629cc9-en

Autor, D., Dorn, D., Katz, L.F., Patterson, C. & Van Reenen, J. (2020). The Fall of the Labor Share and the Rise of Superstar Firms. *The Quarterly Journal of Economics*, 135(2), 645–709. https://doi.org/10.1093/qje/qjaa004

Baqaee, D.R. & Farhi, E. (2020). Productivity and Misallocation in General Equilibrium. *The Quarterly Journal of Economics*, 135(1), 105–163. https://doi.org/10.1093/qje/qjz030

Bauer, A., Garbinti, B. & Georges-Kot, S. (2018). Financial Constraints and Self-Employment in France, 1945-2014. Insee, *Document de travail* N° G018/08. https://www.insee.fr/en/statistiques/3640484

Barkai, S. (2020). Declining Labor and Capital Shares, *Journal of Finance,* 75(5), 2421–2463. https://doi.org/10.1111/jofi.12909

Berlingieri, G., Blanchenay, P. & Criscuolo, C. (2017), The great divergence(s). *OECD Science, Technology and Industry Policy Papers* N°39, OECD Editions, Paris, https://doi.org/10.1787/953f3853-en.

Blundell, R. & Bond, S.(2000). GMM Estimation with persistent panel data: an application to production functions. *Econometric Reviews*, 19(3), 321–340, 2000. https://doi.org/10.1080/07474930008800475

Bonfiglioli, A., Crinò, R & Gancia, G. (2019). Concentration in International Markets: Evidence from US Imports. CEPR *Discussion Paper* N°13566. https://ssrn.com/abstract=3346357

Boussard, J. & Lee, R. (2020). Competition, Profit Share and Concentration. Insee, *Document de travail* N°G2020/04. https://www.insee.fr/en/statistiques/4651858

Caliendo, L., Monte, F. & Rossi-Hansberg, E. A. (2015). The Anatomy of French Production Hierarchies. *Journal of Political Economy*, 123(4), 1–75. https://doi.org/10.1086/681641

Card, D., Heining, J. & Kline, P. (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality. *The Quarterly Journal of Economics*, 128(3), 967–1015. https://doi.org/10.1093/qje/qjt006

CEA (2016). Benefits of competition and indicators of market power. Council of Economic Advisers, *Issue Brief*, 2016.
https://obamawhitehouse.archives.gov/sites/default/files/page/files/20160414_cea_competition_issue_brief.pdf

Cette, G., Koehl, L. & Philippon, T. (2019). The Labor Schare in the Long Term: A Decline? Economie et Statistique / Economics and Statistics, 510-511-512, 35–51. https://doi.org/10.24187/ecostat.2019.510t.1993

De Loecker, J., Eeckhout, J. & Unger, G. (2020). The Rise of Market Power and the Macroeconomic Implications. *The Quarterly Journal of Economics*, 135(2), 561–644. https://doi.org/10.1093/qje/qjz041

De Loecker, J. & Warzynski, F. (2012). Markups and Firm-Level Export Status. *American Economic Review*, 102(6), 2437–2471. https://doi.org/10.1257/aer.102.6.2437

Decker, R. A., Haltiwanger, J., Jarmin, R. S. & Miranda, J. (2016a). Where has all the skewness gone? The decline in high-growth (young) firms in the U.S, *European Economic Review*, Elsevier, vol. 86(C), pages 4–23. https://doi.org/10.1016/j.euroecorev.2015.12.013

Decker, R. A., Haltiwanger, J., Jarmin, R. S. & Miranda, J. (2016b). The Secular Decline in Business Dynamism in the U. S. *American Economic Review*: Papers & Proceedings, 106(5), 203–207.
http://dx.doi.org/10.1257/aer.p20161050

Decker, R. A., Haltiwanger, J., Jarmin, R. S. & Miranda, J. (2016c). Declining Business Dynamism: What We Know and the Way Forward. *American Economic Review,* 106 (5), 203–07.
http://doi.org/10.1257/aer.p20161050

Deroyon, J. (2015). De nouvelles données issues du profilage des groupes. *In:* Insee Références, *Les entreprises en France*, 39–52. https://www.insee.fr/fr/statistiques/1906511?sommaire=1906539

Di Giovanni, J., Levchenko, A. A. & Méjean, I. (2014). Firms, Destinations, and Aggregate Fluctuations. *Econometrica*, 82(4), 1303–1340. https://doi.org/10.3982/ECTA11041

Doraszelski, U. & Jaumandreu, J. (2019). Using Cost Minimization to Estimate Markups.
https://faculty.wharton.upenn.edu/wp-content/uploads/2019/07/robust_markups20190625.pdf

Elsby, M. W. L., Hobijn, B. & Sahin, A. (2013). The decline of the US labor share. *Brookings Papers on Economic Activity*, Fall(2), 1–63.
https://www.brookings.edu/wp-content/uploads/2016/07/2013b_elsby_labor_share.pdf

Garbinti, B., Goupille-Lebret, J. & Piketty, T. (2018). Income inequality in France, 1900-2014: Evidence from Distributional National Accounts (DINA). *Journal of Public Economics*, 162(C), 63–77.
https://doi.org/10.1016/j.jpubeco.2018.01.012

Gutiérrez, G. & Philippon, T. (2017). Investmentless growth: An empirical investigation. *Brookings Papers on Economic Activity*, Fall(1), 89–190.
https://www.brookings.edu/wp-content/uploads/2018/02/gutierreztextfa17bpea.pdf

Gutiérrez, G. & Philippon, T. (2018). How EU Markets Became Free: A Study of Institutional Drift. National Bureau of Economic Research, *Working Paper* N°24700. https://www.nber.org/papers/w24700

Grassi, B. (2017). IO in I-O: Competition and Volatility in Input-Output Networks.
https://sites.google.com/site/grassibasile/home/research/job-market-paper

Grossman, G. M., Helpman, E., Oberfield, E. & Sampson, T. (2018). The Productivity Slowdown and the Declining Labor Share. Society for Economic Dynamics, *2018 Meeting Paper* N°169.
https://economicdynamics.org/meetpapers/2018/paper_169.pdf

Hopenhayn, H. A. (1992). Entry, Exit, and firm Dynamics in Long Run Equilibrium. *Econometrica*, 60(5), 1127–1150. https://doi.org/10.2307/2951541

Karabarbounis, L. & Neiman, B. (2014). The Global Decline of the Labor Share. *The Quarterly Journal of Economics*, 129(1), 61–103. https://doi.org/10.1093/qje/qjt032

Karabarbounis, L. & Neiman, B. (2018). Accounting for Factorless Income. National Bureau of Economic Research, *Working Paper* N°24404. https://www.nber.org/papers/w24404

Karahan, F. Pugsley, B. & Sahin, A. (2019). Demographic Origins of the Startup Deficit. Federal Reserve Bank of New York, *Staff Reports* N°888.
https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr888.pdf

Kehrig, M. & Vincent, N. (2018). The Micro-Level Anatomy of the Labor Share Decline. National Bureau of Economic Research, *Working Paper* N°25275. https://www.nber.org/papers/w25275

Lashkari, D., Bauer, A. & Boussard, J. (2019). Information Technology and Returns to Scale. Banque de France *Working Paper* N°737, 2019. https://publications.banque-france.fr/en/information-technology-and-returns-scale

Levinsohn, J., Petrin, A. & Levinsohn, J. (2003). Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies*, 70(2), 317–341. https://doi.org/10.1111/1467-937X.00246

Mayer, T., Melitz, M. J. & Ottaviano, G. I. P. (2014). Market size, competition, and the product mix of exporters. *American Economic Review*, 104(2), 495–536. https://doi.org/10.1257/aer.104.2.495

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, 71(6), 1695–1725. https://doi.org/10.1111/1468-0262.00467

Olley, G. S. & Pakes, A. (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 64(6), 1263–1297. https://doi.org/10.2307/2171831

Panon, L. (2020). Labor Share, Foreign Demand and Superstar Exporters. https://ludovicrwpanon.github.io/ludovicpanon/JMP_Panon.pdf

Piketty, T., Saez, E. & Zucman, G. (2018). Distributional national accounts: Methods and estimates for the United States. *The Quarterly Journal of Economics*, 133(2), 553–609. https://doi.org/10.1093/qje/qjx043

Song, J, Price, D. J., Guvenen, F., Bloom, N. & von Wachter, T. (2018). Firming Up Inequality. *The Quarterly Journal of Economics*, 134(1), 1–50. https://doi.org/10.1093/qje/qjy025

Van Ark, B. & Jäger, K. (2017). Recent Trends in Europe's Output and Productivity Growth Performance at the Sector Level, 2002-2015. *International Productivity Monitor*, 33(1), 8–23. http://www.csls.ca/ipm/33/vanArk_Jager.pdf

Van Reenen, J. (2018). Increasing Differences between firms: Market Power and the Macro-Economy. Center for Economic Performance, *Working Paper* N°1576 http://cep.lse.ac.uk/pubs/download/dp1576.pdf

DATA

Industry Codes

Industry classification has changed over the 1985-2016 period. From 1985 to 1993 the classification applicable was the NAP (*Nomenclature d'activités et de produits*). It changed to NAF (*Nomenclature d'activités française*) in 1993, since then revised twice (NAF rév. 1 in 2003, NAF rév. 2 in 2008). There is no one-to-one correspondence between these classifications. As a result we make the choice to map each NAP industry code to its most often associated NAF industry code. Similarly we map each NAF industry code to its most often associated NAF rév. 1 industry code, and each NAF rév. 1 code to its most often associated NAF rév. 2. As a result we are able to associate to each firm for each year its industry code in the NAF rév. 2 classification.

Variable Definitions

Our data provide information on total sales of goods, services and merchandises, as well as variations in inventory and immobilized production. For inputs, they provide the book value of tangible and intangible capital, the wage bill and payroll taxes, and the cost of materials, merchandise, and other intermediary inputs. All data on sales, cost of inventory variations and cost of inputs are recorded separately for merchandise and other inputs. We follow definitions from the National Accounts and define output as the sum of immobilized production, variations in inventory, and sales excluding the cost of merchandise; and we define intermediary inputs use as the sum of material expenditures minus inventory variations, and other external inputs. These definitions mean that gross output includes the net margin on merchandise sold, not gross sales of merchandise. Importantly, our data also include the cost of purchased external services in intermediary inputs. Our micro data are in current prices, and we do not observe firm-level prices of intermediary and capital inputs, nor output prices. We deflate nominal values of gross output, intermediary inputs, and capital stock at the NA38 sectors level using price indexes for investment output, and value added from the September 2018 release of the Insee Annual National Accounts.

Data Cleaning

We exclude micro-firms and profiled enterprises from the 2008-2016 data. Firm-year observations of very high or negative labor shares that stem from very low or negative value added observations relative to the firm average across years are replaced with the average labor share of the firm across years. Concentration measures are computed using sales on the entire sample of firms, labor share decomposition and all subsequent analysis are conducted on the sample of firms with at least one salaried employee. The parameters of the translog production function are estimated using a smaller sample of firms with sales above 1 million, and positive value added, intermediary inputs and capital. We also exclude from the estimation sample firms with wage, labor productivity, or capital per employee in the top or bottom 0.1%.

APPENDIX 2

DECOMPOSITION

The decomposition method presented below is applied to aggregate labor share and aggregate inverse markups.

Industry Level Decomposition

Let $k \in \{1, \cdots, K\}$ be some industry classification (e.g., 3 digits in micro data), M stands for an aggregate measure (labor share or markup). Also, let S_k and M_k stand respectively for the weight of the industry in total value added or total sales, and the industry average measure. Define for any variable X:

$$\Delta X_t \equiv X_t - X_{t-1},\ \ \bar{X}_t \equiv \frac{1}{2}\left(X_t + X_{t-1}\right)$$

$$\Delta_T X \equiv X_T - X_0$$

where T is the last period and 0 is the first period. Our first decomposition is:[(i)]

$$\Delta_T M \equiv \underbrace{\sum_{t=1}^{T}\sum_{k} \bar{S}_{kt}\Delta M_{kt}}_{\text{within industries}} + \underbrace{\sum_{t=1}^{T}\sum_{k} \Delta S_{kt}\ \bar{M}_{kt}}_{\text{across industries}} \qquad \text{(B.1)}$$

This allows us to distinguish the extent to which the aggregate variation in markup or labor share is due to a change of industry shares or a within industry variation, irrespective of the sectoral composition of the economy.

Within Industry Decomposition

Next, we focus on changes in the industry-level measure. Our aim is to decompose the changes at the industry level to the changes in the distribution of firm level markup or labor share and the changes in the markup or labor share for the firms of a given quantile. Let $y \in \left[\underline{y};\ \bar{y}\right]$ denote a given level of the labor share or markup. We can write the industry-level outcome as:

$$M_{kt} \equiv \int_{\underline{y}}^{\bar{y}} S_{kt}\left(y\right)\ M_{kt}\left(y\right)\ dy \qquad \text{(B.2)}$$

where $S_{kt}\left(y\right)$ is the density function. In a discrete version, $S_{kt}\left(y\right)$ is the market shares of firms in industry k with labor share or markup close to y, and $M_{kt}\left(y\right)$ denotes the weighted average outcome (labor share or markup) of firms with outcome close to y in industry k at time t. We can now decompose[(ii)]

$$\Delta M_{kt} = \underbrace{\int_{\underline{y}}^{\bar{y}} \bar{S}_{kt}\left(y\right)\Delta M_{kt}\left(y\right)dy}_{\text{within quantiles}} + \underbrace{\int_{\underline{y}}^{\bar{y}} \Delta S_{kt}\left(y\right)\bar{M}_{kt}\left(y\right)dy}_{\text{across quantiles}} \qquad \text{(B.3)}$$

We now summarize the within-industry component change in aggregate measure into the following components:

1. The across quantiles component: $\sum_{t=1}^{T}\sum_{k}\bar{S}_{kt}\int_{\underline{y}}^{\bar{y}}\Delta S_{kt}\left(y\right)\bar{M}_{kt}\left(y\right)dy$
2. The within quantiles component: $\sum_{t=1}^{T}\sum_{k}\bar{S}_{kt}\int_{\underline{y}}^{\bar{y}}\bar{S}_{kt}\left(y\right)\Delta M_{kt}\left(y\right)dy$

(i) This is simply because: $\Delta\left(S_t M_t\right) = \bar{S}_t \Delta M_t + \Delta S_t \bar{M}_t$ *and*

$$\Delta_T\left(SM\right) = \sum_{t=1}^{T}\Delta\left(S_t M_t\right)$$

(ii) As emphasized by Kehrig & Vincent (2018) this decomposition is conceptually distinct from standard within and cross firm decompositions. Let Ω_{kt} *be the set of firms active in time* t*, and* $\bar{\Omega}_{kt}$ *be the set of firms common between time* t *and* $t-1$*,* Ω_{kt}^{+} *the set of new firms at time* t*, and* Ω_{kt}^{-} *the set of firms exiting between time* t *and* $t+1$*. We can then write:*

$$\Delta M_{kt} \equiv \underbrace{\sum_{i\in\Omega_t^-}\bar{S}_{it}\Delta M_{it}}_{\text{within firms}} + \underbrace{\sum_{i\in\Omega_{kt}^-}\Delta S_{it}\bar{M}_{it}}_{\text{across firms}} + \underbrace{\left(\sum_{i\in\Omega_{kt}^+}S_{it}M_{it} - \sum_{i\in\Omega_{kt-1}^-}S_{it-1}M_{it-1}\right)}_{\text{net entry}}$$

where again shares are computed within the industry.

LABOR SHARE, MARKUP, AND TECHNOLOGY

In a first exercise, we do not isolate the contribution of reallocation to the aggregate labor share and write the weighted average mean for a given variable Z:

$$\mathbb{E}_t^{RF}[Z] \equiv \sum_i S_{it} Z_{ti} \tag{C.1}$$

where RF stands for "representative firm".

In a second exercise, we take into account the contribution of reallocation and write the unweighted average mean for a given variable Z :

$$\mathbb{E}_t^{WR}[Z] \equiv \frac{1}{N_t} \sum_i Z_{it} \tag{C.2}$$

where N_t is the total number of firms and WR stands for "with reallocation".

Equation (8) can be rewritten using the definition in equation (C.1), which gives a decomposition of the aggregate labor share into the markup, labor intensity and returns to scale of the representative firm:

$$\Lambda_t = \mathbb{E}_t^{RF}\left[\alpha\gamma\mu^{-1}\right] = \mathbb{E}_t^{RF}[\alpha] \times \mathbb{E}_t^{RF}[\gamma] \times \mathbb{E}_t^{RF}\left[\mu^{-1}\right] + COV_t^{RF} \tag{C.3}$$

or using the definition in equation (C.2), which gives a decomposition of the aggregate labor share into a reallocation term, defined by the gap between weighted and unweighted average labor share, and firm-level unweighted average markups, labor intensity and returns to scale:

$$\Lambda_t = \left(\mathbb{E}_t^{RF}\left[\alpha\gamma\mu^{-1}\right] - \mathbb{E}_t^{WR}\left[\alpha\gamma\mu^{-1}\right]\right) + \mathbb{E}_t^{WR}[\alpha] \times \mathbb{E}_t^{WR}[\gamma]: \\ \times \mathbb{E}_t^{WR}\left[\mu^{-1}\right] + COV_t^{WR} \tag{C.4}$$

where in both cases COV_t^R, gathers all of the covariance terms. This term is positive when firms that have high levels of labor intensity also have high returns to scale and low markups. For each $R \in (RF, WR)$, this quantity is defined by:

$$COV_t^R = cov_t^R\left(\alpha,\gamma,\mu^{-1}\right) + \mathbb{E}_t^R[\alpha]\, cov_t^R\left(\gamma,\mu^{-1}\right) \\ + \mathbb{E}_t^R[\gamma]\, cov_t^R\left(\alpha,\mu^{-1}\right) + \mathbb{E}_t^R\left[\mu^{-1}\right] cov_t^{RF}(\alpha,\gamma)$$

where for all set of variables $(X^s)_{s\in S}$:

$$cov_t^{RF}\left((X^s)_{s\in S}\right) = \mathbb{E}_t^{RF}\left[\prod_{s\in S}\left(X_t^s - \mathbb{E}_t^{RF}\left[X^s\right]\right)\right]$$

Defining as above $\overline{X_t}$ and $\Delta X_t = (X_t - X_{t-1})$ as:

$$\overline{X_t} = \frac{1}{2}(X_t + X_{t-1}), \Delta X_t = (X_t - X_{t-1})$$

we can decompose the variation of the product of expectations in equations (C.3) and (C.4) into contributions of the variation in automation, returns to scale and markups:

$$\Delta\mathbb{E}_t^R[\alpha] \times \mathbb{E}_t^R[\gamma] \times \mathbb{E}_t^R\left[\mu^{-1}\right] = \tag{C.5}$$

$$= \underbrace{\frac{\Delta\mathbb{E}_t^R[\alpha]}{3}\left(\overline{\mathbb{E}_t^R[\gamma] \times \mathbb{E}_t\left[\mu^{-1}\right]} + 2\overline{\mathbb{E}_t^R[\gamma]} \times \overline{\mathbb{E}_t^R\left[\mu^{-1}\right]}\right)}_{\text{Contribution of labor intensity}}$$

$$+ \underbrace{\frac{\Delta\mathbb{E}_t^R[\gamma]}{3}\left(\overline{\mathbb{E}_t^R[\alpha] \times \mathbb{E}_t^R\left[\mu^{-1}\right]} + 2\overline{\mathbb{E}_t^R[\alpha]} \times \overline{\mathbb{E}_t^R\left[\mu^{-1}\right]}\right)}_{\text{Contribution of returns to scale}}$$

$$+ \underbrace{\frac{\Delta\mathbb{E}_t^R\left[\mu^{-1}\right]}{3}\left(\overline{\mathbb{E}_t^R[\alpha] \times \mathbb{E}_t^R[\gamma]} + 2\overline{\mathbb{E}_t^R[\alpha]} \times \overline{\mathbb{E}_t^R[\gamma]}\right)}_{\text{Contribution of markups}}$$

for $R \in (RF, WR)$. By adding to the decomposition in equation (C.5) the variation of the covariance term and of the reallocation term if $R = WR$, we obtain the decomposition of the variation of the aggregate labor share $\Delta\Lambda_t$.

Economie et Statistique / Economics and Statistics

Objectifs généraux de la revue

Economie et Statistique / Economics and Statistics publie des articles traitant de tous les phénomènes économiques et sociaux, au niveau micro ou macro, s'appuyant sur les données de la statistique publique ou d'autres sources. Une attention particulière est portée à la qualité de la démarche statistique et à la rigueur des concepts mobilisés dans l'analyse. Pour répondre aux objectifs de la revue, les principaux messages des articles et leurs limites éventuelles doivent être formulés dans des termes accessibles à un public qui n'est pas nécessairement spécialiste du sujet de l'article.

Soumissions

Les propositions d'articles, en français ou en anglais, doivent être adressées à la rédaction de la revue (redaction-ecostat@insee.fr), de préférence en format MS-Word. Il doit s'agir de travaux originaux, qui ne sont pas soumis en parallèle à une autre revue. Un article standard fait environ 11 000 mots en français (y compris encadrés, tableaux, figures, annexes et bibliographie, non compris éventuelles annexes en ligne). Aucune proposition initiale de plus de 12 500 mots (11 500 pour les soumissions en anglais) ne sera examinée.

La soumission doit comporter deux fichiers distincts :

- Un fichier d'une page indiquant : le titre de l'article ; le prénom et nom, les affiliations (maximum deux), l'adresse e-mail et postale de chaque auteur ; un résumé de 160 mots maximum (soit environ 1 050 signes espaces compris) qui doit présenter très brièvement la problématique, indiquer la source et donner les principaux axes et conclusions de la recherche ; les codes JEL et quelques mots-clés ; d'éventuels remerciements.
- Un fichier anonymisé du manuscrit complet (texte, illustrations, bibliographie, éventuelles annexes) indiquant en première page uniquement le titre, le résumé, les codes JEL et les mots-clés.

Les propositions retenues sont évaluées par deux à trois rapporteurs (procédure en « double-aveugle »). Les articles acceptés pour publication devront être mis en forme suivant les consignes aux auteurs (accessibles sur https://www.insee.fr/fr/information/2410168). Ils pourront faire l'objet d'un travail éditorial visant à améliorer leur lisibilité et leur présentation formelle.

Publication

Les articles sont publiés en français dans l'édition papier et simultanément en français et en anglais dans l'édition électronique. Celle-ci est disponible, en accès libre, sur le site de l'Insee, le jour même de la publication ; cette mise en ligne immédiate et gratuite donne aux articles une grande visibilité. La revue est par ailleurs accessible sur le portail francophone Persée, et référencée sur le site international Repec et dans la base EconLit.

Main objectives of the journal

Economie et Statistique / Economics and Statistics publishes articles covering any micro- or macro- economic or sociological topic, either using data from public statistics or other sources. Particular attention is paid to rigor in the statistical approach and clarity in the concepts and analyses. In order to meet the journal aims, the main conclusions of the articles, as well as possible limitations, should be written to be accessible to an audience not necessarily specialist of the topic.

Submissions

Manuscripts can be submitted either in French or in English; they should be sent to the editorial team (redaction-ecostat@insee.fr), preferably in MS-Word format. The manuscript must be original work and not submitted at the same time to any other journal. The standard length of an article is of about 10,000 words (including boxes if needed, tables and figures, appendices, bibliography, but not counting online appendices if any). Manuscripts of more than 11,500 words will not be considered. Submissions must include two separate files:

- A one-page file providing: the title of the article; the first name, name, affiliation-s (at most two), e-mail et postal addresses of each author; an abstract of maximum 160 words (about 1050 characters including spaces), briefly presenting the question(s), data and methodology, and the main conclusions; JEL codes and a few keywords; acknowledgements.
- An anonymised manuscript (including the main text, illustrations, bibliography and appendices if any), mentioning only the title, abstract, JEL codes and keywords on the front page.

Proposals that meet the journal objectives are reviewed by two to three referees ("double-blind" review). The articles accepted for publication will have to be presented according to the guidelines for authors (available at https://www.insee.fr/en/information/2591257). They may be subject to editorial work aimed at improving their readability and formal presentation.

Publication

The articles are published in French in the printed edition, and simultaneously in French and in English in the online edition. The online issue is available, in open access, on the Insee website the day of its publication; this immediate and free online availability gives the articles a high visibility. The journal is also available online on the French portal Persée, and indexed in Repec and EconLit.

N° 520-521 - 2020

Economie et Statistique

Economics and Statistics

Prochain numéro / Next issue

Varia

ISBN 978-2-11-151277-1 - ISSN 0336-1454 - ECO 520-521
Parution décembre 2020 - PRIX : 17,20 €