French spatial inequalities in an historical perspective

Florian Bonnet

Université Paris 1 Panthéon-Sorbonne Ecole doctorale 465 - Economie Panthéon-Sorbonne

Thèse

présentée pour obtenir le titre de DOCTEUR en
Economie

French Spatial Inequalities in an Historical Perspective

Florian Bonnet

Soutenue publiquement le 11/12/2018 devant un jury composé de :

Rapporteurs	Ryan Edwards, Research Asociate	Berkeley Population Center
	Farid Toubal, Professeur des Universités	ENS Paris-Saclay
Examinateurs	Magali Barbieri, Chargée de Recherche	INED & UC Berkeley
	Lionel Kesztenbaum, Chargé de Recherche	INED & PSE
	Josselin Thuilliez, Chargé de Recherche	CNRS & CES
Directeur de thèse	Hippolyte d'Albis, Directeur de Recherche	CNRS & PjSE

Paris Jourdan Sciences Economiques

Abstract

This Ph. D. thesis has a dual purpose. First, it presents the methods used to build two new historical databases relating to départements. The first database provides the departmental lifetables for the period 1901-2014. The second database provides the departmental distributions of income over the period 1960-2014. Second, this thesis presents the first works resulting from the joint use of these two databases and other statistics: they concern both the dynamics of spatial inequalities and some specific historical events. Thus, the analysis of the spatial distribution of the population since the middle of the 19th century allows to understand the dynamics induced by the rural exodus, but also by the new trends of today's migrations. The analysis of mortality inequalities over the last 200 years shows that inequalities have fallen dramatically since the end of the 19th century, while the geography of excessmortality has changed. Finally, the analysis of spatial income inequalities reveals a continuous decline since the 1920s. This decline occurred only since 1950 if spatial inequalities are observed using asynthetic indicator of welfare, combining both mortality inequalities and income inequalities. The thesis ends with the analysis of internal migrations during the Second World War: these migrations were massive, and clearly oriented towards the free zone. These results testify both to the impact of this event on French demography, and to the quest for freedom of the French of that time, little hampered by the demarcation line.

Keywords

French Spatial Inequalities, Quantitative History, Mortality Inequalities, Income Inequalities, Internal Migrations, Departmental Lifetables.

Les inégalités spatiales en France : une analyse historique

Résumé

Cette thèse a un double objectif. En premier lieu, elle présente les méthodes ayant permis de construire deux bases de données historiques relatives aux départements français. La première met à disposition les tables de mortalité départementales sur la période 1901-2014. La seconde permet de disposer des distributions départementales de revenu sur la période 1960-2014. En second lieu, cette thèse présente les travaux issus de l'utilisation conjointe de ces deux bases de données et d'autres statistiques : ils concernent aussi bien les dynamiques longues des inégalités spatiales que certains évènements historiques. Ainsi, l'analyse de la répartition spatiale de la population depuis le milieu du 19ème siècle permet de comprendre à la fois la dynamique induite par l'exode rural, mais aussi par les nouvelles tendances des migrations d'aujourd'hui, différentes selon les âges. L'analyse des inégalités de mortalité depuis 200 ans montre quant à elle que les inégalités ont largement baissé depuis la fin du 19ème siècle, alors que la géographie de la surmortalité a profondément changé. Enfin, l'analyse des inégalités spatiales de revenus révèle une baisse continue des inégalités depuis les années 1920, baisse qui n'intervient que depuis 1950 si l'on introduit les inégalités de mortalité dans un indicateur synthétique de bien-être. La thèse se conclut par l'analyse des migrations internes durant la seconde guerre mondiale : leur caractère à la fois massif et à destination de la zone libre témoigne aussi bien de l'impact qu'a eu cet évènement sur la démographie française que de la formidable quête de la liberté des français de l'époque, peu entravée par la ligne de démarcation.

Mots-clés

Inégalités spatiales en France, Histoire quantitative, Inégalités de mortalité, Inégalités de revenus, Migrations internes, Tables de mortalité départementales.

Remerciements

Tous mes remerciements et ma gratitude vont dans un premier temps à Hippolyte d'Albis qui, audelà de la direction de cette thèse, a réussi à me convaincre de rentrer de plein pied dans le monde de la recherche. La relation que nous avons initiée au début de l'année 2014 dans le cadre de mon mémoire s'est poursuivie par la suite. C'est une relation à la fois de confiance et d'exigence qui nous a permis de construire ce projet de recherche passionnant aux ramifications nombreuses. Son soutien moral et logistique aura été déterminant, tout comme son idée de la recherche, à la confluence de différents champs disciplinaires. Cette thèse, conjonction d'économie, de démographie, d'histoire et de géographie lui doit énormément.

Je remercie également les membres du jury d'avoir accepté cette tâche. En premier lieu, je voudrais remercier Farid Toubal d'avoir accepté la charge d'être rapporteur. Son expertise et son regard fin sur les techniques de la recherche, ainsi que les discussions que nous avons eues à l'ENS Paris-Saclay durant ces trois dernières années, m'auront permis d'améliorer ma pratique scientifique, et de m'ouvrir certaines perspectives auxquelles je n'aurais pas pensé seul.

En second lieu je suis reconnaissant à Ryan Edwards d'avoir accepté lui aussi la charge d'être rapporteur. Ses travaux à la frontière de l'économie et de la démographie, et en particulier les liens entre incertitude sur la durée de vie et bien-être, auront été pour moi une vraie source d'inspiration. Ses remarques auront par ailleurs permis à ce manuscrit d'être meilleur.

Je voudrais également remercier Magali Barbieri, qui a suivi de très près mes travaux de recherche durant ces trois années. Son expertise dans la construction de la *Human Mortality Database* m'a permis d'initier le projet de la construction de la base de données départementale portant sur la démographie française. Nos nombreux échanges concernant la méthodologie à mettre en oeuvre, ainsi que ses remarques sur les résultats qui en sont issus m'ont permis de me sentir pleinement encadré sur des sujets pour lesquels ma formation initiale était lacunaire.

Ma gratitude va également à Lionel Kesztenbaum. Ses recherches en histoire démographique lui ont permis de procéder à une relecture particulièrement instructive de mes travaux, et notamment ceux portant sur les inégalités de mortalité et sur l'exode durant la seconde guerre mondiale. Par ailleurs, sa disponibilité et ses cours à l'Ecole d'Economie de Paris auront été pour moi un véritable avantage au début de ma thèse.

Je voudrais enfin remercier Josselin Thuilliez d'avoir accepté de faire partie de ce jury. Ses travaux en économie de la santé, et plus particulièrement en épidémiologie, sont pour moi une source d'inspiration dans la perspective future d'étendre la base de données créée aux causes de décès.

Mes travaux de recherche ont été particulièrement influencés par ceux de Thomas Piketty. La lecture attentive de ses écrits est une source d'inspiration inégalable, et la rigueur de sa démarche scientifique un exemple constant à suivre. J'ai été très honoré de pouvoir échanger avec lui et de profiter de ses conseils dans le cadre de cette thèse. Nos différentes rencontres à la rentrée 2017 auront permis de lui donner un nouveau souffle.

Je voudrais également remercier les structures qui m'ont accueilli durant ma thèse. J'ai eu l'occasion de travailler dans des conditions idéales à l'Ecole Normale Supérieure Paris-Saclay, que ce soit dans le département Economie-Gestion ou dans l'antenne du Centre d'Economie de la Sorbonne. Un grand merci notamment à François Pannequin et Sabine Sépari grâce auxquels j'ai pu effectuer ma mission d'enseignement dans cet établissement, à Nicolas Drouhin pour les nombreuses discussions passionnantes que nous avons eues, à Emmanuelle Taugourdeau pour sa bienvaillance et à Jean-Christophe Tavanti pour ses nombreux conseils durant l'année 2018. Je tiens également à remercier l'Ecole d'Economie de Paris dans son ensemble, qui m'a permis de rencontrer un grand nombre de chercheurs de renommée mondiale. Par ailleurs, la qualité de son soutien logistique et la réactivité de son pesonnel administratif ne sont plus à démontrer.

Une thèse ne serait pas une thèse si elle ne permettait pas des échanges passionnés entre doctorants sur les différents champs de recherche, la beauté et la difficulté de la tâche, ainsi que les perspectives d'avenir. Je voudrais donc remercier tout particulièrement Julien Monardo, Samuel Ligonnière, Morgane le Breton, Marine Salès, Maeva Ropaul, Elsa Perdrix, Sébastien Laffitte et Aurélie Sotura, qui m'ont permis de moins ressentir la "solitude du thésard". Leurs conseils, relectures, échanges et soutiens auront été particulièrement précieux.

Au-delà de la communauté scientifique, je voudrais adresser mes remerciements les plus chaleureux à toutes les personnes qui m'ont accueilli durant mes longues heures passées dans les archives. Mes pensées vont ainsi aux bibliothécaires de l'INSEE, et en particulier à Thierry Couderc, dont la réactivité n'a d'égale que sa connaissance des fonds de cette institution. Elles vont aussi aux documentalistes du Centre des Archives Economiques et Financières à Savigny-le-Temple, et en particulier à Marianne Ségaud qui aura réussi à trouver des documents administratifs dont personne ne connaissait encore l'existence.

Pour terminer, je voudrais remercier ma famille pour leur soutien sans faille depuis de très longues années, et en particulier mes parents, ma soeur et mon grand-père. L'éloignement n'a jamais distendu les liens qui nous unissent.

Contents

List of Figures

List of Tables

Introduction

Dans son livre La Grande Evasion, santé, richesse et origine des inégalités (2016), Angus Deaton, prix de la banque de Suède en mémoire d'Alfred Nobel 2015, pointe du doigt l'extraordinnaire amélioration des conditions de vie sur Terre depuis maintenant plus de deux siècles : "La vie est aujourd'hui meilleure qu'à aucune autre époque de l'histoire . Il y a plus de gens plus riches, et moins de gens vivant dans une pauvreté atroce. Nous vivons plus longtemps, et les parents n'ont plus pour habitude de voir mourir un sur quatre de leurs enfants". Il rappelle ainsi en ce qui concerne la mortalité "qu'un enfant de sexe féminin qui naît aujourd'hui aux Etats-Unis peut espérer vivre 80 ans. [...] C'est un changement remarquable par rapport à la situation de son arrière grand-mère, née en 1910, qui avait à sa naissance une espérance de vie de 54 ans".

L'étude de l'évolution du niveau de vie en termes de revenu et de mortalité est au coeur de l'analyse de Deaton, et en ce sens se démarque des analyses menées généralement, qui se concentrent exclusivement sur le niveau de vie en termes de revenu. Cette analyse est particulièrement pertinente lorsque l'on s'intéresse au bien-être d'un individu. En effet, une analyse rapide du revenu de cycle de vie permet de montrer que l'amélioration du bien-être peut passer par deux facteurs différents. Le premier est l'augmentation de son revenu pour une période précise de sa vie : un individu vivra dans des conditions préférables si son revenu augmente de 10%. On peut qualifier cette amélioration du bien-être d'amélioration intensive. Le second facteur est l'amélioration de sa durée de vie : à revenu égal pour toutes les périodes de sa vie, un individu préferera vivre 80 ans que 70 ans. On peut qualifier cette amélioration du bien-être d'amélioration extensive.

Ainsi, un des résultats essentiels du livre de Deaton (2016) est de montrer que cette Grande Evasion a permis aux habitants de nombreux pays développés de vivre plus riches, plus longtemps, et en plus grand nombre, et que ce mouvement se perpétue de nos jours. Son analyse, qui est faite au niveau international, ne permet pourtant pas de cerner les évolutions particulières de la France. Cette première partie de l'introduction dresse donc un tableau national de l'évolution de la population, ainsi que des conditions de vie en termes de revenu et de mortalité depuis 200 ans, en mettant en perspective ces évolutions françaises avec les évolutions constatées dans des pays tels que l'Allemagne, le Royaume-Uni, les Etats-Unis ou encore l'Italie.

La Grande Evasion en France depuis 200 ans

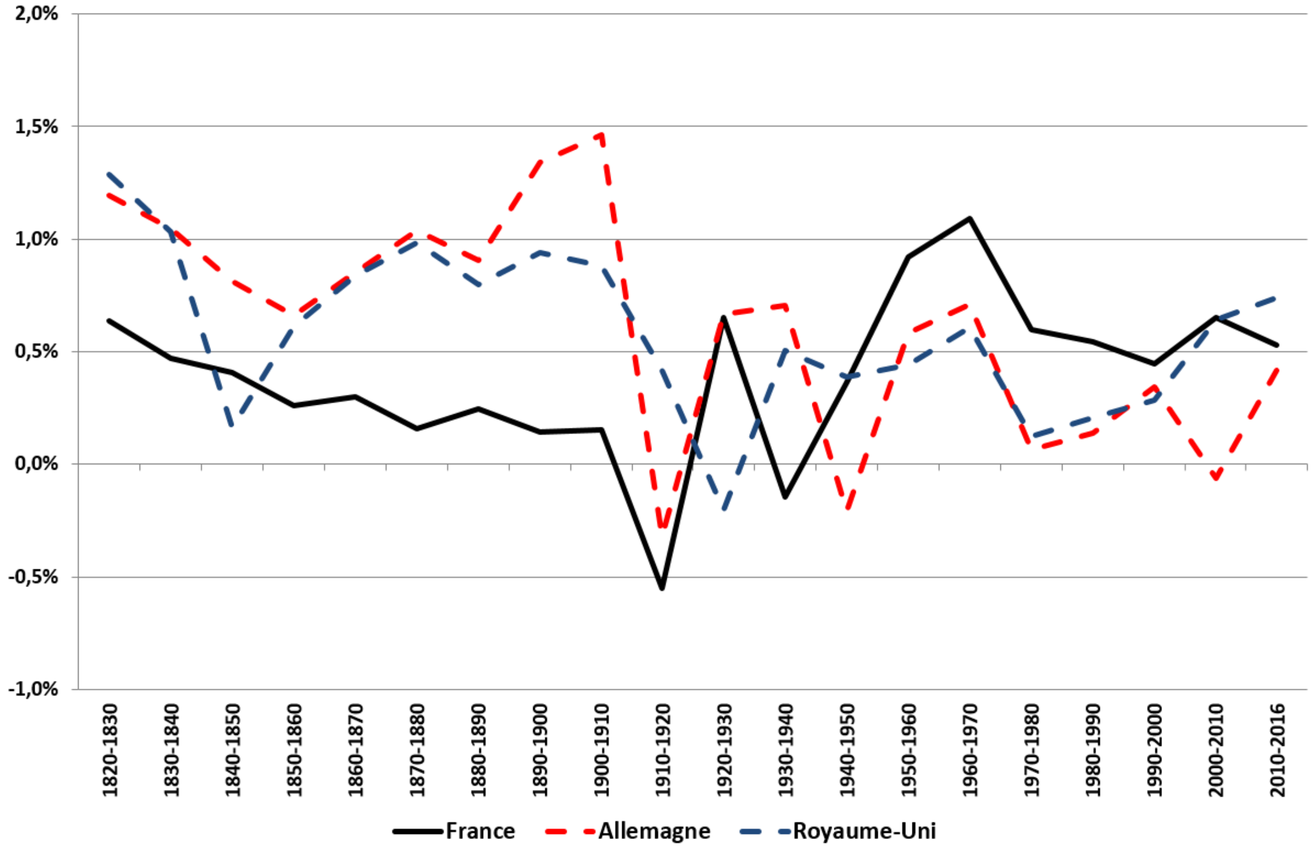
Les évolutions de population en France et dans le monde

Le 19ème siècle : la France au défi d'une croissance atone de sa population

Grâce aux statistiques établies par Bolt et al. (2018), il est possible d'analyser l'évolution de la population des plus grands pays dans leurs frontières actuelles, et ce depuis le début du 19ème siècle. Le Tableau 0.1 présente ces évolutions de population pour la France et 4 autres grands pays occidentaux (Allemagne, Royaume-Uni, Italie, Etats-Unis) entre 1820 et 2016, et ce pour certaines dates clés.

Le premier constat est qu'en 1820, la France était le pays le plus peuplé du monde occidental. Sa population était de 31 250 000 individus, alors que celle de l'Italie était de 20 176 000, celle de l'Allemagne de 24 905 000, celle du Royaume-Uni de 21 239 000 et enfin celle des Etats-Unis de 9 981 000. Elle représentait environ 30% de la population totale des 5 pays cités précédemment. Par la suite, le 19èmesiècle a constitué une période de changement majeur en ce qui concerne l'importance de la France sur la scène européenne et internationale. Pour le comprendre, j'ai représenté sur la Figure 0.0.1 le taux de croissance annuel moyen de la population nationale, et ce pour chaque décennie. J'ai également représenté sur cette figure le taux de croissance annuel moyen de la population du Royaume-Uni et de l'Allemagne, les deux grands rivaux de la France à cette époque.

Figure 0.0.1 : Taux de croissance annuels moyens de la population en France, au Royaume-Uni et en Allemagne, 1820–2016



Sources : Maddison Project Database, Bolt et al. (2018).

Entre 1820 et 1910, le taux de croissance de la population française est positif, mais à un niveau assez faible : il était de 0,6% au début du 19ème siècle, et de 0,2% au début du 20ème siècle. En ce qui concerne le Royaume-Uni et l'Allemagne, les taux de croissance de la population étaient quant à eux à des niveaux plus élevés, aux alentours de 1% en moyenne. Cette différence, qui peut sembler limitée à première vue, est lourde de conséquences : à la veille de la première guerre mondiale, la France comptait un peu plus de 41 millions d'habitants, alors que l'Allemagne en comptait près de 63 millions et le Royaume-Uni près de 45 millions. La part de la France dans la population des 5 pays cités précédemment a largement chuté, pour atteindre seulement 15% en 1910, alors que celle de l'Allemagne était restée stable aux alentours de 23%. Dans un monde où les guerres se gagnent en fonction de la taille des contingents militaires, ce constat a posé un vrai problème politique largement relayé à l'époque. Les causes de ce manque de dynamisme de la population sont aujourd'hui assez bien documentées. La France a connu une transition démographique très précoce, et le coefficient transitionnel associé (c'est-à-dire le facteur multiplicatif de la population avant et après la transition démographique) a été très faible : taux de mortalité et taux de natalité ont baissé de manière quasi simultanée.

Le 20ème siècle : au-delà des guerres, un dynamisme de la population retrouvé

Les deux guerres mondiales ont constitué un traumatisme démographique majeur pour la France, l'Allemagne et le Royaume-Uni. Les taux de croissance moyens de la population sont devenus négatifs pour la première fois, et la population a stagné de manière globale : la France n'a gagné que 1 300 000 habitants entre 1910 et 1950, l'Allemagne 5 500 000. Les pertes militaires ont créé de forts déséquilibres, notamment entre les sexes, avec une proportion de femmes dans la population totale en forte augmentation.

Néanmoins, la période postérieure à la seconde guerre mondiale a montré un dynamisme nouveau en termes de croissance de la population, et ce dans tous les pays de panel. Ce dynamisme a été particulièrement fort en France : la Figure 0.0.1 montre que le taux de croissance moyen de la population a été supérieur à 1% entre 1950 et 1970, et aux alentours de 0,6% entre 1970 et nos jours. La part de la population française dans la population totale des pays du panel a stagné aux alentours de 12%, quand celle de l'Allemagne, de l'Italie et du Royaume-Uni s'est mise à décroître. Alors que la population de l'Italie et du Royaume-Uni avait surpassé celle de la France durant la première moitié du 20ème siècle, la population de la France est à nouveau supérieure à celle de ces pays en 2016. Ainsi, à cette date, la France compte près de 67 millions d'habitants, contre seulement 60,7 millions en Italie et 66 millions au Royaume-Uni.

TABLE 0.1 : POPULATION EN FRANCE ET DANS LE MONDE, 1820–2016

	1820	1850	1880	1910	1950	1980	2016
France	31 250	36 350	39 045	41 224	42 518	55 110	66 957
Italie	20 176	24 460	29 534	36 572	47 105	56 451	60 738
Allemagne	24 905	33 746	43 500	62 884	68 375	78 298	83 707
Royaume-Uni	21 239	27 181	34 623	44 916	50 127	56 314	65 888
Etats-Unis	9 981	23 580	50 458	92 767	152 271	227 726	324 656

Sources : Maddison Project Database, Bolt et al. (2018).

Les évolutions de la production par habitant en France et dans le monde

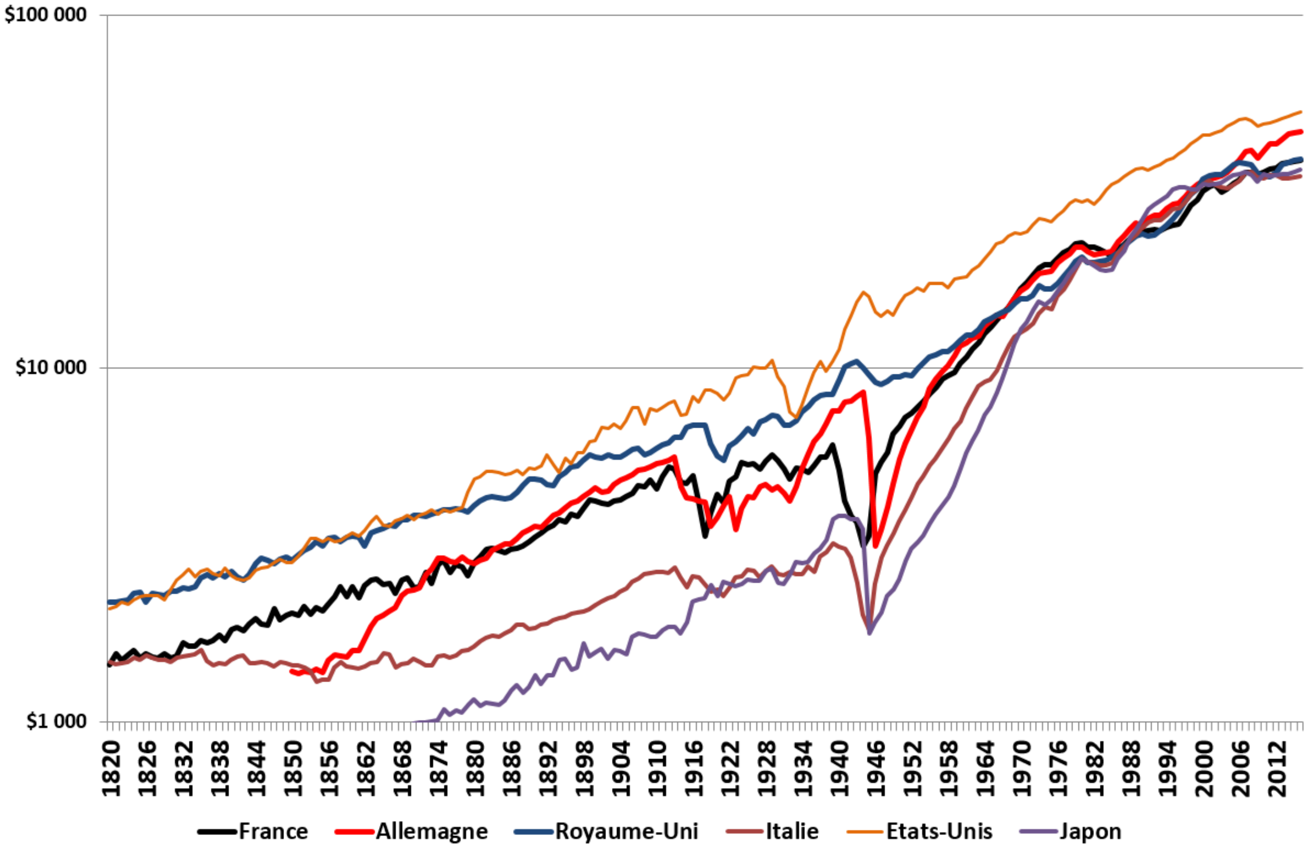
Le 19ème siècle : Révolution industrielle et lent processus de convergence

Les données de Bolt et al. (2018) permettent également de disposer du Produit Intérieur Brut d'un grand nombre de pays dans le monde (en dollars américains de 2011), et ce depuis le début ou le milieu du 19ème siècle selon les pays. Disposant également de la population, il est alors possible de calculer les PIB par habitant d'un certain nombre de pays et d'analyser leur évolution sur cette période. La Figure 0.0.2 représente ainsi le PIB par habitant des 5 pays déjà suivis précédemment, auxquels s'ajoute le Japon à partir du milieu du 19ème siècle. On peut voir que la France disposait en 1820 d'un PIB par habitant d'environ 1 400 dollars, valeur qui n'a cessé d'augmenter entre 1820 et 1910, à un rythme moyen d'environ 1,7% par an. La croissance la plus forte sur cette période a été observée entre 1890 et 1900, au coeur de la période que les historiens appellent aujourd'hui la "Belle Epoque". La Révolution Industrielle, qui a transformé progressivement un pays agricole en une puissance industrielle, a conduit la France à rattraper progressivement son retard sur le Royaume-Uni, et dans une moindre mesure sur les Etats-Unis. En 1910, l'écart entre le PIB par habitant français et le PIB par habitant anglais n'était plus que de 30%, alors qu'il était de 50% en 1820. Le même constat peut être fait pour l'Allemagne, dont la croissance de la production par habitant a été particulièrement rapide entre 1860 et 1910, aux alentours de 2,5% par an. Alors que cette production par habitant était inférieure de 30% à celle de la France en 1850, elle était de 10% supérieure en 1910. Déclassée par rapport à son vieil ennemi au regard de son poids dans la population totale à la veille de la Première Guerre Mondiale, la France l'était aussi en ce qui concerne la richesse produite par habitant.

Le 20ème siècle : destructions puis croissance inégalée

La première moitié du 20ème siècle fût une période de grande incertitude et de crises. Entre 1910 et 1945, le PIB par habitant a fluctué, impacté à la fois par les deux guerres mondiales et la crise de 1929 : en 1945, il était inférieur de 25% à sa valeur de 1910, et d'environ 3 300$. L'évolution fût similaire en Italie et au Japon, mais aussi en Allemagne, fortement touchée durant l'entre-deux-guerres. Le Royaume-Uni et les Etats-Unis n'ont pas connu ce ralentissement, enregistrant des taux de croissance moyens d'environ 2% sur la période.

Figure 0.0.2 : PIB par habitant en France et dans le monde, 1820–2016



Sources : Maddison Project Database, Bolt et al. (2018).
Notes : PIB par habitant en dollars américains de 2011.

La deuxième moitié du 20ème siècle est quant à elle une période d'expansion. Toutes les économies sans exception ont connu une progression très forte de leur production par habitant durant ce que les historiens ont appelé "les 30 glorieuses". En France, le taux de croissance entre 1950 et 1980 était d'environ 4,5%, similaire à celui de l'Allemagne. Ces taux étaient encore plus élevés en Italie et au Japon (respectivement de 6,5% et 7%), et plus faibles au Royaume-Uni et aux Etats-Unis (aux alentours de 2,5%). Après 1980, ces taux de croissance ont faibli pour s'établir aux alentours d'un peu moins de 2% de manière générale, sous l'effet à la fois des chocs pétroliers et de la diminution des gains de productivité. Au final, en 2016, la production par habitant en France est 27 fois plus élevée que ce qu'elle était en 1820.

En ce qui concerne les inégalités entre pays, les évolutions différenciées de la période 1950–1980 ont conduit à une convergence des niveaux de production par habitant : au début des années 1980, ces derniers sont tous compris entre 19 000 et 22 000$, à l'exception des Etats-Unis qui continuent à mener la marche (30 000$ en 1981). Sur la période récente, les économies semblent à nouveau diverger, l'Allemagne rattrapant progressivement son retard sur les Etats-Unis alors que la France, le Royaume-Uni, l'Italie et le Japon semblent aujourd'hui plus en retrait.

Les évolutions de l'espérance de vie à la naissance en France et dans le monde

Le 19ème siècle : une augmentation tardive de l'espérance de vie

Les données de la Human Mortality Database permettent enfin de disposer des espérances de vie à la naissance d'un certain nombre de pays, pour des périodes plus ou moins longues. La Figure 0.0.3 présente ces espérances de vie à la naissance pour les 5 pays déjà étudiés précédemment, ainsi que pour la Suède, pour laquelle les données historiques sont les plus fiables et proposent le recul le plus long. Les espérances de vie à la naissance qui sont présentées sont calculées par période : c'est donc la durée de vie qu'un nouveau-né peut espérer atteindre si les conditions de mortalité du moment restaient les mêmes tout au long de sa vie. Elles sont calculées pour les deux sexes confondus.

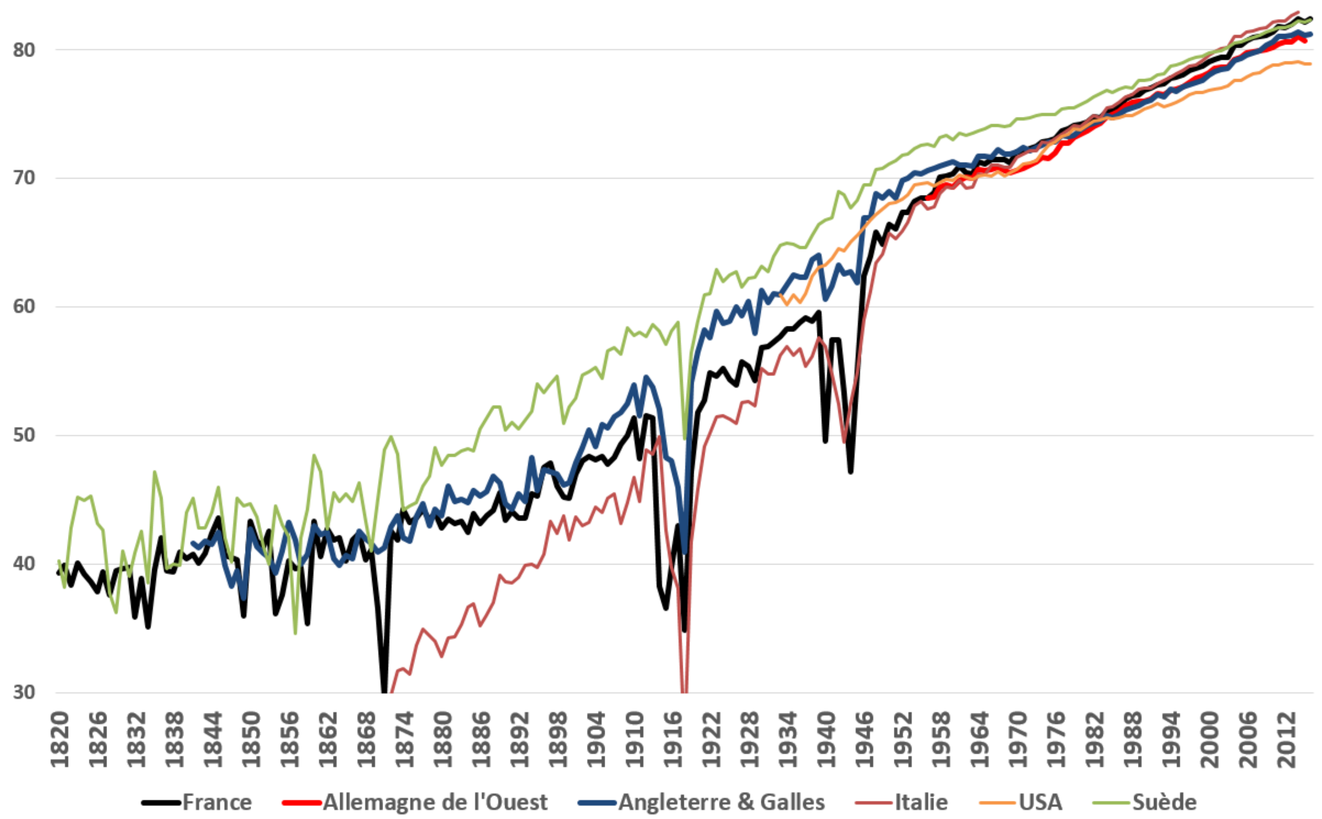
Durant une très large première moitié du 19ème siècle, l'espérance de vie à la naissance en France n'a pas augmenté : ainsi, entre 1820 et 1869, celle-ci est restée aux alentours de 40 ans. Par ailleurs, cette espérance de vie à la naissance a connu de nombreuses fluctuations du fait des épidémies qui ont continué à toucher de manière périodique le pays. C'est notamment le cas des épidémies de choléra, qui ont entraîné des pics importants de mortalité dans les années 1830 et 1850. Ainsi, entre 1853 et 1854, l'espérance de vie à la naissance a diminué de plus de 6 années, passant de 42,5 à 36,2 ans. Si l'on s'intéresse aux autres pays pour lesquels les statistiques sont disponibles pour la même période, on peut voir que la Suède a connu elle aussi de larges fluctuations de son espérance de vie, légèrement supérieure à l'espérance de vie française. En Angleterre et au Pays de Galles, les fluctuations ont été moins fortes, mais l'espérance de vie a suivi sensiblement la même évolution qu'en France.

Après la guerre contre la Prusse, qui a entraîné une baisse très forte de l'espérance de vie en France (elle s'établissait en 1871 à moins de 30 ans), celle-ci a amorcé un mouvement de hausse sans précédent jusqu'à la Première Guerre Mondiale : en 1914, elle était de quasiment 50 ans, soit une augmentation d'environ un trimestre en moyenne par an. Cette hausse fût sensiblement moins forte que celle observée en Italie, ce qui peut s'expliquer par la forte convergence des conditions de mortalité pour ce pays. Elle fût également moins forte que celle observée en Suède, dont la hausse sur la même période s'est établie à 13 années. Elle fût néanmoins similaire à celle observée au Royaume-Uni. Par ailleurs, on peut voir que les fluctuations annuelles étaient moins marquées qu'auparavant, même si certains évènements restés dans la mémoire collective sont visibles sur la Figure 0.0.3. C'est le cas par exemple de la canicule de 1911, qui marque un creux juste avant la Première Guerre Mondiale.

Le 20ème siècle : des désastres des deux guerres à la hausse continue de l'espérance de vie

La première moitié du 20ème siècle fût globalement une période de hausse de l'espérance de vie à la naissance en France : celle-ci est passée de 50 ans en 1914 à 62 ans en 1946, soit sensiblement le même rythme de croissance que la période 1872–1914. Cependant, cette période est marquée par de fortes fluctuations, du fait des crises majeures qui ont touché le pays. C'est d'abord la Première Guerre Mondiale, durant laquelle l'espérance de vie a atteint un plancher de 36,5 ans en 1915. Cette baisse fût également observée en Italie, ainsi que dans une moindre mesure en Angleterre et au Pays de Galles. Ce conflit, particulièrement meurtrier, a causé la mort de plus 1 300 000 militaires en France. La seconde crise est liée à l'épidémie de grippe espagnole, qui décima en 1918 et 1919 une large partie

Figure 0.0.3 : Espérance de vie à la naissance en France et dans le monde, 1820–2016



Sources : Human Mortality Database (2018).

de la population européenne déjà fortement touchée par les 4 années de guerre. En France, ce sont près de 250 000 personnes qui en seraient mortes. Durant cette crise, tous les pays furent touchés, ce qui n'avait pas été le cas de la Suède durant le conflit. Dans ce pays, l'espérance de vie à la naissance a diminué de 9 ans pour s'établir à moins de 50 ans. Enfin, la Seconde Guerre Mondiale constitue la troisième crise : la Figure 0.0.3 montre qu'elle a fortement impacté la France et l'Italie, dont les espérances de vie ont à nouveau diminué de près de 10 années entre 1939 et 1944. La Suède, les Etats-Unis n'ont pas connu ces pics de mortalité dus au conflit. De manière générale, la France sort de cette période avec une espérance de vie à la naissance plus faible que la Suède et les Etats-Unis, mais aussi que l'Angleterre et le Pays de Galles, alors que les valeurs étaient sensiblement les mêmes en 1914.

Enfin, la période 1946–2016 fût une période de hausse de l'espérance de vie à la fois ininterrompue et faite à un rythme encore jamais connu : elle est passée de 62 à 82 ans en France, soit un rythme de 3,5 mois gagnés par an. Par ailleurs, les inégalités d'espérance de vie entre les différents pays se sont considérablement réduites durant cette période. En 2014, elles étaient toutes situées entre 81 et 83 ans. Une exception notable concerne les Etats-Unis, dont l'espérance de vie a sensiblement moins augmenté que les autres pays entre 1970 et 2014 : en fin de période, elle s'établissait à seulement 79 ans. Elle a par la suite légèrement décliné, laissant craindre un impact sur cet indicateur de la crise des opioïdes qui sévit dans ce pays. Si l'on s'intéresse au classement des différents pays, il est intéressant de noter que les positions relatives ont changé : la Suède, qui a longtemps mené le peloton, avait en 2014 une espérance de vie inférieure à celle de la France et de l'Italie. A l'inverse, l'Italie a longtemps été en retard mais est aujourd'hui pris en exemple. Enfin, l'Angleterre et le Pays de Galles ont longtemps profité d'une espérance de vie supérieure à celle de la France au 20$^{\text{ème}}$siècle mais ont été dépassé dans le milieu des années 1970.

La statistique au niveau local

Dans son livre, Deaton montre que la Grande Evasion n'a pas profité à tous les pays de ce monde, et que certains sont aujourd'hui encore coincés dans des trappes à pauvreté et à forte mortalité. La révolution industrielle et la transition épidémiologique, en permettant aux pays aujourd'hui développés de voir leurs conditions de vie s'améliorer à un rythme sans précédent, a aussi entraîné une hausse des inégalités au niveau international, inégalités qui sont encore aujourd'hui à un niveau élevé.

Deaton se concentre sur les inégalités internationales, sans rien dire des inégalités intranationales. Or, si les grandes tendances historiques présentées précédemment permettent de comprendre les évolutions de population, de revenu (approximé par la production par habitant) et de mortalité de la France aux 19$^{\text{ème}}$ et 20$^{\text{ème}}$ siècles, elles ne disent rien de ces évolutions au niveau local. Pourtant, il est fort probable que ces évolutions n'aient pas été les mêmes sur tout le territoire. En ce qui concerne l'évolution de la population, on peut prendre deux exemples bien connus des démographes pour illustrer la question. Le premier est l'exode rural, qui a conduit à la diminution de la population dans les zones peu densément peuplées au bénéfice des villes. Le deuxième est l'apparition de la "diagonale du vide" durant le 20$^{\text{ème}}$ siècle : cette zone peu densément peuplée, qui s'étend des frontières de la Belgique à celles de l'Espagne en passant par le Massif Central, a vu sa population stagner voire diminuer depuis la fin de la seconde guerre mondiale, quand celle de la France dans son ensemble augmentait fortement du fait de la hausse de la natalité. Un autre exemple bien connu illustre les évolutions différenciées du revenu selon les territoires : le Nord-Est de la France, qui a fondé sa richesse sur ses ressources naturelles et la sidérurgie, est aujourd'hui fortement touché par la concurrence internationale. La reconversion de son appareil productif se fait difficilement, ce qui entraîne chômage et diminution de la population.

Le but de cette thèse a donc été de répliquer cette analyse de l'évolution des conditions de vie menée par Deaton au niveau international, mais dans un cadre national, celui de la France. En lieu et place d'une étude des évolutions différenciées selon les pays, j'ai analysé sur longue période les évolutions différenciées de population, de revenu, et de mortalité au niveau local. Pour cela, j'ai exploité deux caractéristiques françaises majeures qui font de ce pays un objet d'étude idéal pour procéder à une analyse fine des évolutions historiques au niveau local. La première est le découpage du territoire national en départements, dont les limites géographiques sont restées relativement stables. Le deuxième est l'existence depuis deux siècles d'un appareil statistique efficace, qui permet de disposer d'une quantité importante de données brutes à exploiter.

La longue histoire des départements français métropolitains

Avant la Révolution Française, le territoire métropolitain est décrit comme un ensemble de découpages administratifs (militaire, ècclésiastique, fiscal...) ne se recoupant pas, et dont la lisibilité pour les contemporains était très faible. Ces derniers se référaient encore aux anciennes provinces, caractérisées par leur patrimoine culturel et linguistique propre : on parlait alors de la Touraine, du Béarn, du Roussillon ou encore de la Flandre.

Avec la Révolution Nationale, un premier projet de réorganisation administrative fût proposé par l'abbé Sieyès à l'Assemblée Nationale le 7 Septembre 1789. Il prévoyait le découpage du territoire national en 81 départements carrés, de 18 lieues de côté chacun (à l'exception de Paris). Ce projet fût finalement abandonné au profit d'un second, basé sur les spécificités hydrographiques et géologiques locales, et s'inscrivant dans l'ensemble dans le cadre des anciennes provinces. Par exemple, la Franche-Comté y était divisée en trois départements (le Doubs, le Jura et la Haute-Saône), le Dauphiné en trois départements (les Hautes-Alpes, la Drôme et l'Isère). De plus, chaque département devait être créé de manière à ce que les citoyens puissent se rendre dans leur chef-lieu en moins d'une journée de cheval. Les 83 nouvelles unités géographiques créées prirent effet le 4 mars de l'année 1790. Leur nombre a largement fluctué entre cette date et 1815, du fait notamment des guerres napoléoniennes. Ainsi, ils étaient 130 en 1811 avec les annexions territoriales en Allemagne, aux Pays-Bas, en Belgique, en Italie, en Suisse et en Espagne[1]. En 1815, les limites administratives furent à nouveau celles de 1790, à l'exception de la création des départements du Vaucluse, du Tarn-et-Garonne et de la scission du département de Rhône-et-Loire en deux départements, le Rhône et la Loire. A ces 86 entités furent ajoutés les départements des Alpes-Maritimes, de la Savoie et de la Haute-Savoie en 1860, puis celui du Territoire de Belfort en 1871 à la suite de la guerre contre la Prusse. Les 90 départements sont restés stables jusqu'en 1968, date à laquelle la réorganisation de l'Ile-de-France a porté leur nombre à 95. Enfin, en 1976, deux départements corses, la Corse-du-Sud et la Haute-Corse, furent créés pour remplacer l'ancien département unique. Ces deux départements ont laissé place à une collectivité unique en 2018.

Cette brève histoire des départements métropolitains montre bien que leurs limites administratives sont restées relativement stables. Or cette stabilité permet à ces entités géographiques d'être suivies dans le temps, sans que ne se posent de gros problèmes liés aux changements territoriaux. Nous verrons dans cette thèse que le peu de modifications observées a entraîné certaines complications. Il était donc essentiel que ces dernières ne soient pas trop nombreuses, et ce afin de procéder à une analyse historique sur longue période qui ne soit pas biaisée.

Des statistiques démographiques et économiques disponibles au niveau départemental

Le deuxième atout, l'existence d'un appareil statistique efficace au niveau départemental, est consubstantiel de la forte présence de l'Etat dans les départements. En effet, leur création en 1790 par l'As-

[1] Le territoire national s'étendait alors jusqu'aux rives du Rhin et aux bouches de l'Elbe au Nord, incluait Bâle à l'Est, descendait jusqu'à Rome au Sud-Est et incorporait notamment Barcelone et Tarragone au Sud-Ouest.

semblée Constituante a essentiellement permis à l'Etat de mieux administrer le territoire national. La création des préfets et des préfectures par Napoléon Bonaparte en 1800, dans le but de ramener une paix pérenne à la suite des évènements de la Révolution, allait dans ce sens. Aujourd'hui, les préfets sont nommés par décret du Président de la République, sur proposition du premier ministre. Se devant de garder une neutralité politique absolue et d'appliquer la politique du gouvernement, ils ont *"la charge des intérêts nationaux, du contrôle administratif et du respect des lois"* (article 72 de la Constitution). De manière factuelle, leurs missions consistent à veiller au maintien de l'ordre public, au bon déroulement des élections locales, à la répartition entre les différents échelons locaux des dotations de l'Etat, à la bonne santé environnementale du territoire et enfin à la mise en place des politiques nationales (éducation, santé, emploi, aménagement du territoire). Cette forte présence de l'Etat dans les départements va de pair avec des pouvoirs très limités pour les représentants du peuple à ce niveau administratif. Aujourd'hui, les compétences du conseil départemental sont limitées notamment à l'aide sociale (protection de l'enfance, insertion des personnes en difficulté, aide aux personnes handicapées...), à l'éducation à travers le réseau des collèges, ou encore à la voirie (routes départementales).

Les évolutions de population, de production ou encore de mortalité que nous avons analysées au niveau national sont basées sur des statistiques récupérées par les chercheurs dans des publications officielles qui ont traversé les siècles. Il en va ainsi de la *Statistique du Mouvement de la Population* par exemple, qui permet de disposer de façon annuelle du nombre de naissances et du nombre de décès selon l'âge du défunt. Ces statistiques sont disponibles depuis le début du 19ème siècle. Il en va ainsi également des *Renseignements Statistiques Relatifs aux Impôts Directs*, publication qui permet de disposer de façon annuelle là encore du nombre de foyers fiscaux soumis à l'impôt sur le revenu, classés par tranches de revenu imposable. Ces statistiques sont disponibles depuis la création de l'impôt sur le revenu, intervenue durant la Première Guerre Mondiale.

Or il est essentiel de comprendre que ces statistiques construites au niveau national ne l'ont pas été à partir de bulletins individuels provenant directement des individus, des foyers fiscaux ou encore des communes. En effet, une première centralisation a été effectuée de manière systématique au niveau départemental, avant que ces centralisations départementales ne soient utilisées pour procéder à une centralisation au niveau national. Cette caractéristique fondamentale est particulièrement utile dans le cas qui nous intéresse ici, puisqu'en théorie toutes les études faites au niveau national peuvent être répliquées au niveau départemental.

La longue marche vers la création de bases de données fiables

Recherches en archives, numérisation et digitalisation

Une grande partie des premiers mois de cette thèse a donc consisté à partir à la recherche des statistiques brutes potentiellement disponibles. La première étape fût de recenser les statistiques disponibles dans les publications officielles, similaires à celles dans lesquelles les statistiques nationales furent décelées. En ce qui concerne la démographie au niveau départemental, les données ont été retrouvées en grande partie de cette manière. La *Statistique du Mouvement de la Population* met ainsi à disposition des chercheurs le total des naissances domiciliées ainsi que les décès par âge quinquennaux domiciliés pour chaque département, chaque sexe et chaque année depuis 1855. Par ailleurs, les recensements ont fait l'objet d'un dépouillement exhaustif depuis 1851, dépouillement qui permet pour chacun d'entre eux de disposer de la population par âges quinquennaux pour chaque département et chaque sexe.

La seconde étape consista à recenser les statistiques disponibles dans d'autres documents, non encore connus par la communauté des chercheurs. Cette phase fut primordiale pour reconstruire les données de revenu pour chaque département. En effet, les *Reseignements Statistiques Relatifs aux Impôts Directs* ne permettaient pas de disposer des mêmes statistiques que celles établies au niveau national, en l'occurence les distributions de revenu imposable. Après un long travail de recherche d'indices dans les publications statistiques de l'époque, et notamment celles de l'INSEE, et un grand nombre d'échanges avec les archivistes des différentes places parisiennes, le Centre des Archives Economiques et Financières a pu retrouver des documents administratifs encore inconnus dans lesquels ces statistiques de revenu imposable par tranche étaient disponibles au niveau départemental. Leur existence pour les périodes 1960–1969, 1986–1998 et 2001–2014 seulement m'ont enfin obligé à procéder à de nouvelles recherches dans les publications officielles, et ce afin de reconstruire les revenus moyens départementaux pour les périodes manquantes, en l'occurence 1922–1959, 1970–1985 et 1999–2000.

Une fois ces statistiques identifiées, le travail principal a consisté en leur numérisation et leur digitalisation systématiques. La masse d'informations ainsi recueillie est considérable. A titre d'exemple, la récupération des statistiques des décès par âge et par sexe pour une seule année représente environ 3 600 données uniques[2], celle des statistiques de population par âge et par sexe pour un recensement est de 4 500 données uniques. Au total, et avant la récupération des statistiques d'Etat-Civil auprès de l'INSEE à partir de 1968, intervenue à la fin de l'année 2017, ce sont environ 500 000 données qui ont été ainsi récupérées pour construire la base de données démographique portant sur la population et la mortalité. En ce qui concerne la base de données économique portant sur les distributions départementales de revenu pour les périodes 1960–1969 et 1986–1998, ce sont environ 2 500 tabulations fiscales qui ont été photographiées, soit l'équivalent de 50 000 données uniques digitalisées. Enfin, le calcul des revenus départementaux pour les périodes 1922–1959, 1970–1985 et 2000–2001 a entraîné la numérisation d'environ 25 000 données supplémentaires. Au final, ce sont donc environ 575 000 nombres qui ont été identifiés, photographiés, archivés puis numérisés dans le cadre de cette thèse de doctorat, sans compter environ 25 000 autres qui ont été numérisés sans possibilité d'exploitation à ce jour.

Cette masse de statistiques brutes digitalisées à gérer, qui peut sembler particulièrement importante, ne doit pas faire oublier que ces données ne valent rien si leur fiabilité n'est pas assurée. En effet, ces statistiques ont été récupérées à des périodes différentes, sur des territoires différents, par des organismes différents. Leur qualité peut donc différer spatialement et temporellement, mais aussi selon les variables étudiées. J'ai donc mis en place un certain nombre de vérifications qui me permettent d'affirmer que la qualité des données est suffisante pour analyser les grandes tendances de

[2]Le produit de 90 départements, pour 2 sexes, et 20 groupes d'âge.

l'évolution des inégalités spatiales, comme se propose de le faire cette thèse. Je souhaite ici présenter deux d'entre elles. La première porte sur la cohérence interne des données récupérées. Ainsi, en ce qui concerne la base de données démographique, j'ai vérifié que la somme des décès et des populations par âge était toujours égale au total départemental indiqué dans les tableaux. J'ai également vérifié que les sommes départementales étaient bien égales aux totaux nationaux indiqués eux aussi dans les tableaux. J'ai suivi en cela les préconisations de Bonneuil (1997). Les mêmes vérifications ont été faites en ce qui concerne la base de données économique. La seconde porte sur la cohérence entre les données départementales et les données nationales. Considérant que la qualité des études faites précédemment au niveau national sont aujourd'hui reconnues par la communauté scientifique, il était essentiel de vérifier qu'à minima les sommes départementales coïncidaient avec les valeurs nationales. C'est ce qui a été fait lors de la construction de la base de données démographique, ainsi que lors de la construction de la base de données économique. Il est important de noter que les écarts étaient presque toujours nuls, à l'exception notable des deux guerres mondiales durant lesquelles le total des décès diffère. Cela s'explique par la mobilisation de nouvelles données concernant les décès historiques et militaires.

De la statistique brute à la base de données finale

Une fois ces statistiques brutes identifiées, photographiées, archivées, numérisées et vérifiées, j'ai pu les utiliser pour construire deux bases de données qui sont au coeur de cette thèse. La présentation de la méthodologie est entièrement détaillée dans la Partie II.

A l'intérieur de cette Partie II, le Chapitre 5 porte sur la construction de la base de données démographique. L'objectif de cette base de données est de mettre à disposition de la communauté scientifique les tables de mortalité annuelles par sexe de chaque département de France métropolitaine pour tout le 20$^{\text{ème}}$ siècle. Par ailleurs, elle met également à disposition des estimations annuelles de population par âge et par sexe de chacun de ces départements. Pour arriver à cette fin, le choix de la méthodologie à utiliser était primordial, afin que la qualité des estimations ne soit pas remise en cause. J'ai choisi pour cela le protocole de la *Human Mortality Database* (HMD), développé par des chercheurs de l'Université de Californie (Berkeley, USA), du *Max Planck Institute* (Rostock, Allemagne) et de l'Institut National d'Etudes Démographiques (Paris, France). Ce protocole a permis de reconstruire un grand nombre de tables de mortalité au niveau national sur longue période, tables de mortalité qui ont été reprises dans les meilleurs travaux académiques. J'ai donc dans un premier temps retranscrit le protocole dans le logiciel statistique R, avant de le modifier à la marge afin de tenir compte des spécificités des données départementales pour la période 1901–2014. La première de ces spécificités concerne les décès militaires dus aux deux guerres mondiales, ainsi que les décès en déportation durant la deuxième guerre mondiale. Pour cela, j'ai mobilisé deux sources récentes, la base de données des Morts Pour La France du ministère des armées, et une base de données collaborative regroupant une très large majorité des déportés. La seconde de ces spécificités concerne les changements de limites administratives des départements, dus aux deux guerres mondiales et à la réorganisation de l'Ile-de-France en 1968. Au final, et comme le montre le Chapitre 5, je dispose des tables de mortalité et des populations par âge pour chaque année, chaque sexe et la quasi intégralité des départements pour la période 1901–2014. Les seules données manquantes concernent les départements sous administration allemande entre 1871 et 1918 : les données sont disponibles pour eux à partir de 1921.

Le Chapitre 6, co-écrit avec A. Sotura, porte sur la construction de la base de données économique. L'objectif de cette base de données est de mettre à la disposition des chercheurs les distributions départementales de revenus par foyer fiscal de chaque département de France métropolitaine pour la période 1922–2014. Comme nous l'avons vu précédemment, les statistiques brutes actuellement retrouvées dans les archives ne permettent ces travaux que pour les périodes 1960–1969, 1986–1998 et 2001–2014. Le choix du protocole méthodologique à utiliser était quant à lui naturel, puisque les données brutes à disposition sont exactement les mêmes que celles qui ont été utilisées par Piketty (1997) et Garbinti et al. (2016) pour reconstruire les distributions de revenu au niveau national. Les travaux pionniers de Piketty ont donné naissance à une vaste littérature portant sur l'évolution des inégalités dans un certain nombre de pays développés, et plus récemment sur l'évolution des inégalités dans les pays en développement. Les estimations et analyses qui en sont issues sont aujourd'hui regroupées dans la *World Inequality Database* (WID), et librement accessibles par la communauté scientifique. Là encore, nous avons dans un premier temps retranscrit ce protocole méthodologique dans le logiciel statistique R, avant de le modifier substantiellement pour tenir compte des spécifités départementales. Ces modifications portent quasi exclusivement sur la période 1960–1969, période durant laquelle les statistiques brutes sont de moins bonne qualité. En effet, elles ne concernent que les foyers fiscaux imposables, ce qui nous a obligé à estimer à la fois le nombre de foyers fiscaux total de chaque département à partir des données de démographie, et le revenu total grâce aux estimations de la période 1986–1998. Au final, et comme le montre le Chapitre 6, je dispose des distributions de revenu par foyer fiscal pour chacune des années et chacun des départements de France métropolitaine. Ces estimations sont cohérentes avec les estimations faites par Garbinti et al. (2016) au niveau national.

Par ailleurs, les données recueillies m'ont également permis de reconstruire les revenus moyens départementaux pour les périodes 1922–1959, 1970–1985 et 1999–2000, ce qu'il n'était pas possible de faire à partir des tabulations fiscales. Ce travail ne pouvant être comparé à ceux présentés dans les Chapitres 5 et 6 du fait de son ampleur plus limitée, j'ai choisi de le présenter dans la seconde section du Chapitre 3, puisque c'est exclusivement dans celui-ci que je me suis servi de ces estimations. Il est important de noter qu'il n'y avait pas de protocole méthodologique déjà largement utilisé par la communauté scientifique pour procéder à ces estimations. Je propose donc une méthode spécifique, qui peut être présentée en deux étapes. La première étape consiste à calibrer un modèle économétrique permettant d'estimer les revenus par habitant de chaque département à partir de statistiques démographiques (notamment la structure de la population par âge) et économiques (nombre de foyers fiscaux imposables à l'impôt sur le revenu, revenu des foyers fiscaux imposables soumis à l'impôt sur le revenu, impôt sur le revenu total récolté). Ce modèle économétrique, calibré sur les années 1960–1969, 1986–1998 et 2001–2014 pour lesquelles les estimations sont solides, est utilisé dans un deuxième temps pour imputer les revenus départementaux des années 1922–1959, 1970–1985 et 1999–2000. La clé de la méthodologie est de montrer que le modèle économétrique estime quasi parfaitement les revenus par habitant départementaux des années pour lesquelles je dispose d'estimations fiables. Au final, et comme le montre le Chapitre 3, je dispose des revenus départementaux pour chaque année entre 1922 et 2014.

L'analyse des inégalités spatiales sur longue période

Ces bases de données ont été construites dans le but d'analyser les inégalités spatiales sur longue période dans le cas de la France. En effet, ce champ de recherche est resté assez limité jusqu'à présent. Plusieurs études peuvent néanmoins être citées dans ce domaine. Dans ses travaux portant sur la transition démographique au 19$^{\text{ème}}$ siècle, Bonneuil (1997) survole rapidement la question en montrant que ces inégalités spatiales ont globalement diminué. Le papier de Combes et al. (2011) constitue un autre travail d'importance en ce qui concerne l'analyse sur longue période des inégalités spatiales en France. Les auteurs analysent à la fois la distribution spatiale de la population et de la production totale, ainsi que les inégalités spatiales de production par habitant. Cependant, cette analyse souffre du manque de donnés fiables à utiliser : seules les années 1860, 1930, 1982 et 2000 y sont étudiées. De manière générale, Combes et al. (2011) utilisent l'indice de Theil comme indicateur d'inégalités spatiales et montrent que la répartition de la population, des emplois, des chômeurs et de la production totale était bien plus déséquilibrée en 2000 qu'en 1860. A l'inverse, ils montrent que les inégalités spatiales de production par adulte et de production par emploi ont largement diminué sur la même période. Bazot (2014) enfin, en utilisant de nouvelles données sur la période 1860–1910, affine ces conclusions en montrant que les inégalités spatiales de production par habitant sont restées stables entre 1860 et 1900, avant de diminuer fortement entre 1900 et 1910. D'autres études ponctuelles sont citées dans chacun des chapitres de cette thèse, mais leur lecture amène toutes à la même conclusion : les données disponibles sont relativement limitées, et la méthodologie utilisée ne permet pas de dresser un portrait exhaustif de l'évolution des inégalités spatiales en France. Les bases de données construites et la méthodologie que je propose permet de pallier ces deux limites observées dans la littérature aujourd'hui.

Le Chapitre 1 décrit l'évolution de la distribution spatiale de la population pour la période 1851–2014, en utilisant à la fois la base de données démographique et les données des recensements des années 1851–1896. A première vue, cette question de la distribution spatiale de la population semble éloignée de celle des inégalités spatiales ; pourtant, elle est en au coeur, pour au moins deux raisons. La première est que ma démarche se base non pas sur l'analyse des inégalités entre les territoires, qui auraient tous une importance égale, mais sur l'analyse des inégalités entre les populations qui y habitent. Ainsi, une baisse des inégalités de revenu par habitant n'aura pas la même interprétation en termes de politique publique selon si elle est due à un exode massif de la population des départements les plus pauvres ou à un développement endogène de ces derniers. La seconde raison est que l'analyse de la structure par âge de la population de chaque département permet de poser les premiers jalons d'une explication des inégalités spatiales, notamment en termes de revenus : on peut s'attendre à ce que les départements dans lesquels les personnes âgées sont surreprésentées aient un revenu moyen plus faible que la moyenne nationale. A l'inverse, les départements dans lesquels les personnes en fin de carrière sont surreprésentées devraient avoir un revenu moyen plus élevé que la moyenne nationale, du fait de revenus salariaux plus élevés et d'un taux de chômage plus faible.

Approfondissant l'analyse menée par Combes et al. (2011), je montre notamment dans le Chapitre 1 que la répartition de la population totale est certes de plus en plus déséquilibrée, mais que cette dynamique peut-être décrite comme l'enchaînement de trois phases, que j'appelle *hyper-centralisation, hyper-centralisation contrariée* et *multipolarisation.* Ces trois phases n'auraient pas pu être identifiées sans l'utilisation d'un large panel d'indicateurs d'inégalités : dépassant les simples indices de Gini ou de Theil, qui résument en un indicateur unique la distribution d'une variable (en l'occurence les densités départementales de population), j'utilise 6 indicateurs différents qui permettent de séquencer l'intégralité de cette distribution. Ainsi, dans le cas de la population, étudiée dans ce chapitre 1, je m'intéresse à la part de la population habitant dans les 20% de km^2 les moins densément peuplés, les 10% de km^2 les plus densément peuplés... Dans un deuxième temps, je m'intéresse non plus à la répartition spatiale de la population tous âges confondus, mais à la répartition spatiale de la population pour plusieurs groupes d'âge. Cela me permet de voir le profil des indicateurs d'inégalités selon l'âge, et l'évolution de ce profil dans le temps. Je montre ainsi que ce profil par âge a largement changé depuis 1851. A cette époque, les plus âgés étaient plus inégalitairement répartis que les populations de moins de 60 ans, phénomène s'expliquant par des différences départementales de mortalité. A titre d'exemple, les plus de 60 ans étaient exclus de Bretagne, non pas parce que ces derniers la fuyaient, mais bien parce que les taux de mortalité ne permettaient pas leur survie. Aujourd'hui, ce sont les adultes âgés de 30 à 40 ans qui sont les plus inégalitairement répartis, alors que les plus âgés sont plus égalitairement répartis sur le territoire national. Ce profil particulier pourrait s'expliquer par la localisation des emplois, rassemblés dans les grandes métropoles. Or la localisation des plus âgés n'est pas contrainte par cette localisation des emplois, contrairement aux classes d'âge actif. Enfin, je montre dans ce chapitre que les structures par âge des populations départementales sont de plus en plus différentes, et ce depuis la fin de la seconde guerre mondiale. Certains départements du Sud-Ouest ont une part de jeunes actifs dans leur population bien plus faible que cette même part au niveau national, leur tissu économique reposant essentiellement sur les plus âgés. A l'inverse, les classes d'âge actifs sont fortement surreprésentées dans les départements accueillant les plus grandes métropoles françaises.

Le Chapitre 2, co-écrit avec H. d'Albis, se concentre sur les inégalités spatiales de mortalité sur la période 1806–2014. Pour cela, nous utilisons à la fois les tables de mortalité départementales issues de la base de données démographique pour la période 1901–2014, ainsi que les tables de mortalité départementales calculées par Bonneuil (1997) pour le 19$^{\text{ème}}$ siècle. Ces deux sources nous permettent d'analyser les inégalités spatiales de mortalité chez les femmes sur une période s'étalant sur plus de deux siècles. Dans ce Chapitre 2, nous montrons dans un premier temps que ces inégalités ont dans l'ensemble largement diminué sur la période 1806–2014, et ce quel que soit l'indicateur d'inégalités utilisé. Au-delà de cette tendance de long-terme, la période 1806–1850 voit se succéder des périodes de baisse et de hausse des inégalités, tout comme la période 1980–2014. Dans un deuxième temps, nous avons procédé à une analyse simultanée de l'évolution des inégalités spatiales de mortalité et de l'évolution de l'espérance de vie au niveau national. En effet, il est essentiel de comprendre qu'une baisse des inégalités de mortalité n'a pas la même résonance si elle s'accompagne d'une hausse ou d'une baisse de l'espérance de vie au niveau national. Dans le premier cas, les départements les plus défavorisés rattrapent leur retard, ce qui correspond à une convergence vertueuse. Dans le second cas, les départements les plus favorisés voient leur espérance de vie diminuer, ce qui au contraire n'a rien de vertueux. Grâce à cette analyse simultanée, il est possible de montrer que la France a connu une longue période de convergence vertueuse qui a duré un siècle (1880–1980), raison pour laquelle nous l'avons qualifiée de *convergence centenaire*. Avant cela, le pays a connu une succession de phases de hausses et de baisses de l'espérance de vie, rythmée par les épidémies et les guerres. A partir de 1980, la hausse de l'espérance de vie au niveau national s'est faite en parallèle d'une hausse des inégalités spatiales : les départements les plus favorisés ont ainsi vu leur espérance de vie augmenter plus vite que les autres. Dans un troisième temps, nous avons cartographié l'évolution de la géographie de la mortalité sur deux siècles. Pour cela, nous avons conservé les périodes précédemment identifiées, et regroupé les départements en 3 classes pour chacune d'entre elles. Ainsi, il s'avère que la Bretagne a de tout temps été une zone où la mortalité était plus élevée que la moyenne nationale. Au contraire, le Nord et le Nord-Est de la France, ainsi que le Sud-Est, ont vu leur position relative changer : celle des premiers s'est fortement degradée, alors que le Sud-Est est aujourd'hui une région où l'espérance de vie est supérieure à la moyenne nationale. Ce résultat est vrai également pour les départements urbains, qui souffraient d'une "pénalité urbaine" au 19$^{\text{ème}}$ siècle du fait d'épidémies plus fréquentes et plus dévastatrices.

Le Chapitre 3 se concentre sur les inégalités spatiales de revenus et de bien-être sur la période 1922–2014. Pour cela, j'utilise à la fois les populations et les tables de mortalité départementales issues de la base de données démographique, et les revenus départementaux estimés grâce à la base de données économique. Dans ce Chapitre 3, je montre dans un premier temps que les inégalités de revenu par adulte ont largement diminué sur la période depuis près d'un siècle. Cette diminution s'est faite à un rythme particulièrement élevé entre 1950 et 1980, contrairement aux périodes 1922–1949 et 1980–2014. Je nomme la période 1950–1980 "30 glorieuses des inégalités spatiales" : outre la diminution des inégalités de revenu par adulte, on note durant cette période un rééquilibrage de la distribution spatiale du revenu sur le territoire national. Dans un deuxième temps, je m'intéresse à l'évolution des inégalités spatiales de bien-être. Cette notion de bien-être trouve son origine dans les travaux de Becker et al. (2005). Dans ce papier, les auteurs analysent les inégalités internationales de bien-être, et calculent un revenu équivalent tenant compte à la fois du revenu par habitant et de l'évolution de l'espérance de vie. En d'autres termes, Becker et al. (2005) monétarisent le gain en espérance de vie et ajoutent cette valeur au revenu par habitant. Ayant montré (d'Albis et Bonnet, 2018) que la méthodologie utilisée par Becker et al. (2005) souffre de certaines limites, je calcule le bien-être des adultes vivant dans chaque département en utilisant la méthodologie proposée par Fleurbaey et Gaulier (2009). Au lieu de monétariser les gains en espérance de vie entre deux dates, les auteurs monétarisent les différences d'espérance de vie entre deux pays à chaque date. Ainsi, un pays qui dispose d'une espérance de vie moindre subit une pénalité par rapport au pays qui dispose de l'espérance de vie la plus élevée. L'analyse de l'évolution des inégalités spatiales de bien-être montre que les inégalités spatiales n'ont pas diminué entre 1922 et 1950, mais ont au contraire augmenté.

Cette hausse est due à la corrélation négative entre revenu par adulte et espérance de vie au début du 20$^{\text{ème}}$ siècle, corrélation aujourd'hui positive. En 1922, les départements les plus développés, en l'occurence les départements urbains, souffraient d'une espérance de vie plus faible que la moyenne nationale du fait notamment de l'insalubrité qui y régnait. La convergence des conditions de mortalité montrée au Chapitre 2 a dans un premier temps augmenté les inégalités spatiales de bien-être, puisque les départements urbains ont vu leurs conditions de mortalité se rapprocher de la moyenne nationale. Par la suite, la convergence des revenus par adulte dans les départements les plus pauvres a permis la baisse des inégalités spatiales de bien-être. Dans un troisième temps, en cartographiant la géographie du bien-être et son évolution entre 1922 et 2014, je montre dans ce Chapitre 3 qu'un large quart nord-est est passé d'un niveau de bien-être supérieur à la moyenne nationale en 1922 à un niveau de bien-être inférieur à la moyenne nationale aujourd'hui.

De l'analyse de tendances de long-terme à l'analyse d'évènements spécifiques

Si l'analyse de l'évolution des inégalités spatiales sur très longue période est au coeur de cette thèse, il me semblait important de montrer que les bases de données pouvaient également être utilisées pour analyser des évènements spécifiques et leurs conséquences au niveau local. On pourra penser notamment à l'impact économique et démographique des deux guerres mondiales, de l'épidémie de grippe espagnole en 1919, de la crise de 1929 ou encore du rapatriement des pied-noirs en provenance d'Algérie dans les années 1960. Dans le Chapitre 4, je me concentre plus particulièrement sur les flux de population à l'intérieur de la France durant la Seconde Guerre Mondiale. Si la littérature historique s'est longtemps consacrée à l'analyse de l'exode de Mai-Juin 1940 dû à l'avancée des troupes allemandes, elle n'a jamais cherché à estimer les flux annuels interdépartementaux entre 1939 et 1945. C'est précisément ce que je propose de faire dans ce Chapitre 4. Pour cela, je développe une méthodologie originale qui permet d'estimer les populations de chaque département, et ce pour chaque année d'une période intercensitaire. Cette méthodologie se base sur la connaissance des populations pour chaque année au niveau national, ainsi que des décès par cause et des naissances dans chaque département, ce qui est permis en France par les statistiques de l'Etat-Civil. Connaissant les populations départementales pour chaque année entre 1939 et 1946, il est ensuite facile d'en déduire les flux migratoires. L'analyse de ces flux me permet de montrer que l'exode de Mai-Juin 1940 ne fût qu'une première étape en ce qui concerne les flux massifs de réfugiés observés durant la Seconde Guerre Mondiale. Ainsi, entre 1940 et 1941, une large partie de la population fut déplacée du Nord vers le Sud du pays. Cette analyse révèle également que la ligne de démarcation établie entre le Nord de la France (sous occupation allemande) et le Sud de la France (sous administration française) marque une limite très nette entre les départements qui ont accueilli des réfugés et ceux qui les ont vus partir, et ce jusqu'en 1943. Ainsi, cette ligne de démarcation ne semble pas avoir empêché les réfugiés de transiter du Nord vers le Sud. Enfin, l'analyse des flux de population sur toute la Seconde Guerre Mondiale me permet de montrer que ce conflit a laissé une trace forte dans la démographie départementale : en effet, les réfugiés de certains départements du Nord de la France n'étaient pas encore rentrés en 1946, alors que la population du Sud-Ouest avait augmenté de façon pérenne du fait de ces mouvement migratoires.

Part 1: Essays in French Demographic and Economic History

Chapter 1

Spatial Distribution of Population by Age in France, 1851–2014

Abstract

This paper analyses the uneven spatial distribution of population in the French *départements* and how it evolved between 1851 and 2014. I use a new demographical database built at the departmental level for the 20th century. Firstly, I show that the spatial distribution of the population is increasingly unbalanced. This process can be described as the sequence of three phases called "hyper-centralization", "hyper-centralization thwarted" and "multipolarization". I analyse this process through the two potential sources of population increase, natural movement (total births minus total deaths) and migrations. I point out some geographical regions which cumulate impairments: no attractivity for potential migrations and scarce natural movement. Second, I reveal that the age profile of inequalities has changed along the 150 years. Today, this profile has an inverted U-shape, with a maximal inequality for young workers. Third, I show that population age structures are more and more differentiated between *départements*. The territorial specialization according to population by age has increased since 1950. In 2014, retirees are largely overrepresented in rural *départements*, while 20 to 39-year-old are overrepresented in urban *départements*. As such, the rural South-west excludes more and more young workers.

1.1 Introduction

The issue of regional planning was central in French political debates during the second half of the 20th century. This can be explained by two phenomena at this time: the share of the population living in the Paris region was increasing sharply for 100 years – especially because of the rural exodus – and the share of the population living in the "empty diagonal" – a geographical area connecting *Ardennes* in the North-East to *Ariège* in the South-West (see Oliveau and Doignon, 2016) – was declining. These two phenomena contributed to the popularity of Gravier (1947)'s book entitled "Paris and the French Desert": the author showed how the rural areas gradually emptied in favor of a territorial organization centered around the capital. This question led in 1963 to the creation of the "Delegation for Regional Planning and Regional Attractiveness" (DATAR in French), in charge of implementing the interministerial policy of spatial planning. This policy involved, for example, incentives for companies which settle in depopulated or impoverished territories, or the promotion of "balancing metropolises" in order to reverse the hyper-centralization.

In this paper, I clarify the question of spatial planning by providing an in-depth analysis of the spatial distribution of the French metropolitan population. This study is mainly based on Bonnet (2018b)'s database in which the annual populations by age and *département* since 1901 are available. I add to this database population-by-age data retrieved in censuses between 1851 and 1896 to cover a period of more than 150 years. With these data I analyze the spatial distribution of the population according to age and sex. This question is not clearly understood nowadays. Combes et al. (2011) for example analyzed the spatial distribution of the total population between the French metropolitan *départements* for only five years (1860, 1896, 1930, 1982, 2000). Ayuda et al. (2010) analyzed the evolution of the spatial distribution of total population for 9 European countries including France, but in 1850 and 2000 only. Talandier et al. (2016) propose a cartographic analysis of this issue since 1806 within the framework of French cities, but for the total population and without using an analysis by indicators (see Le Mée, 1989 for a presentation of the raw data used). Finally, other papers have studied this issue in the recent period, without a historical perspective (see in particular Breton et al., 2017).

From a methodological point of view, indicators used to analyze the evolution of the spatial distribution of the population are crucial. So far, the literature has mainly used indicators aggregating the departmental distribution of densities per km^2 into a single indicator. Combes et al. (2011) for example used the Theil indicator, while Ayuda et al. (2010) based their analysis on indicators such as the standard deviation, the coefficient of variation or the Gini index. These indicators may hide evolutions in specific parts of the departmental distribution. For example, the Gini index may decrease while the share of the less populated *départements* decreases. This occurs if this phenomenon is more than offset by a population transfer from the most populated *départements* to the "a-little-less" populated *départements*. In this paper, I try to provide an answer to this issue. In order to analyze the spatial distribution of the population and its evolutions, I use a complete set of indicators containing both the Gini index and indicators specific to each part of the departmental distribution of densities per km^2.

In addition, Bonnet (2018b)'s database provides annual flows of births and deaths by sex and *département*. With these data one can dissociate population variations due to natural or migratory movement. In this paper, I highlight the *départements* combining demographic imbalances: an imbalanced demographic structure that leads to a birth deficit, and a very low attractiveness for migrations. Consequently, I participate in the literature on "Shrinking Regions", which emerged with Oswalt and Reniets (2006)'s work: they listed all the cities in the world whose population decreased over time. This literature analyses the territories too, particularly within the works of Bontje et al. (2012), Fol (2012) or Galjaard et al. (2012). My contribution shed new light on this issue since I do not only look at the absolute variations of population but I also compare it to the national evolution. I therefore consider that a territory is on the decline if its share in the national population decreases, since this evolution leads to a loss of political and economic power.

Finally, with the departmental population structures by age and sex, one can analyze how these structures are increasingly differentiated between the territories and where older or younger people are overrepresented. As far as I know, this issue of differences in age structures has never been treated in the literature but deserves further consideration. Indeed, if age structures are more and more differentiated, the territories are becoming more and more interdependent. This would result in significant transfers of income from the most active to the oldest territories. Here we meet the distinction between productive and residential economies in line with Blanc (2007), Davezies (2008) or Beyers and Nelson (2008). These potential transfers of income implies that fiscal decentralization should be conducted with caution: local budgets must not be affected by demographic imbalances. It also implies that public policies have to be driven by the specificities of each territory: towards education where young people are overrepresented, towards health and dependency where the older are. In this paper, I therefore propose both an overall analysis of differences in age structures by using a single indicator, but also a cartographic analysis to better know territories in wich each age group is overrepresented or underrepresented.

All of these analyzes bring a number of new results. Firstly, I show that the population is more and more unevenly spread. In broad outline, this process can be described with three phases. For example, the increase of inequalities from 1851 to 1901 is the result of the concentration of population in the most densely populated territories to the detriment of all others. Indicators also show that the increase in inequality from 1968 onwards hides a drop in the share of the most densely populated *départements*. Second, I reveal that *départements* which cumulated imbalances according to both natural and migratory movement changed between the first and the second half of the 20th century: they were mainly in the central and western parts of the country between 1901 and 1968, while they were in the North-East and the south of *Massif Central* between 1968 and 2014. As such, I name "wide belt of attractiveness" the *départements* surrounding *Seine* and *Seine-et-Oise* insofar as their population increase is due to both a strong natural movement and migrations. This contrasts with *Seine* and *Seine-et-Oise* whose population growth rate due to migrations is lower than the national rate. Third, the analysis of the spatial distribution of the population according to age shows an inverted U-shaped profile over the recent period: the "20–39" age group is the most unevenly distributed. This profile has changed over time: in the second half of the 19th century, the elderly were the most unevenly distributed. Forth, the analysis of departmental age structures reveals that they are more and more differentiated since the end of the Second World War. This process is mainly explained by an overrepresentation of young adults in urban *départements*, while retirees are overrepresented in rural *départements.* In particular, the rural South West is gradually becoming a land of exclusion for young workers.

The rest of the paper is organized as follows. In Section 2 I present the data as well as the methods used in this study. In Section 3 I present the results. The fourth section concludes.

1.2 Data and Methods

In this paper I analyze the evolutions of the spatial distribution of the population between the French *départements* since 1851. The choice of this geographical unit is explained by the stability of their administrative boundaries since their creation in 1789. For the purpose of this study, I will confine the results to French metropolitan *départements.* Overseas *départements* are not included because available data are too recent to carry out a long-term analysis. In this study the term "national" refers to these French metropolitan *départements.*

In order to analyze the evolution of the spatial distribution of the population, I use differences in population densities, in line with Ayuda et al. (2010) for example. The population density per km^2 is defined as the ratio of the population to the total of km^2. I use population densities since *département*s are not of equal size. For example, *Gironde* in the South-West has an area of 10,375 km^2 while the one of its neighbor *Tarn-et-Garonne* is only 3,718 km^2.

In stage 1, I present the density differences for the total population. My analysis, however, goes further than most of the current works on the spatial distribution of population over long periods: I analyze in stage 2 inequalities for each major age group (0–19, 20–29, 30–39, 40–49, 50–65, 65–79, 80 and over). In the remaining of the paper, I analyze the evolution of the spatial distribution of the female population, even if I have populations for both sexes. This choice is explained by two main reasons. The first concerns the data available to recontextualize the evolutions. Lifetables are crucial to do so : thanks to Bonneuil (1997)'s work, I have female lifetables for the 1851-1900 period, but not male lifetables. The second reason concerns the readibility of the historical trends. France has experienced three major wars during this period, and men have been more widely affected than women by their consequences: forced migration and excess mortality make long-term developments less readable. Nevertheless, when there are noticeable differences, I present the results for the men in the appendix.

1.2.1 Inequality Indicators

There are a large number of inequality indicators to capture inequalities. These indicators can also be used to analyze the spatial distribution of the population. Mackenbach and Kunst (1997) in the field of health studies, or Cowell (2011) more generally, make a non-exhaustive list of these indicators; they show how each of them provides different informations on the issue. One can use the indices based on extreme ranks – the difference or the ratio between the highest value and the lowest value –

or on the interquantile interval (the difference or the ratio between the $x\,\%$ of the higher values and the $(1-x)\,\%$ of the lowest values). There are also Gini or Theil indices that reduce the distribution in a single indicator, or the indices of dissimilarity which express the part that should be distributed among the observations so that the values are similar for all. One may analyze the evolution of inequalities through $\sigma-$convergence or $\beta-$convergence (Barro & Sala-i-Martin (1992) for the most known) . The $\sigma-$convergence studies the evolution of an inequality indicator between two periods. There is $\beta-$convergence if the relationship between the variation of a variable between two dates t_0 and t_1 and the values in t_0 is positive. Finally, inequalities can be analyzed in absolute or relative terms: a density difference of 10 inhabitants per km^2 between two territories represents 20% of the average density when the mean is equal to 50, but only 10% for a value of 100.

In what follows, I use the Gini index as it is easily readable. In order to deepen the analysis, I use however other indicators which target specific parts of the departmental population distribution. I split this distribution in six parts and calculate the share of each of them. I get the shares of the 10% most densely populated km^2 in the national population (namely P90–100), but also the shares of the second (P80–90), third and forth (P60–80), fifth and sixth (P40–60), seventh and eighth (P20–40), ninth and tenth (P0–20).

With these indicators I cannot study the inter-departmental differences in population distribution according to age. To do so, I use the Kullback-Leibler Divergence (Kullback and Leibler, 1951) as d'Albis et al. (2014) did to analyze the international dissimilarities of age-specific mortality rates. This Kullback-Leibler Divergence (KLD) is based on Shannon's entropy (1948). Formally, the KLD between two population distributions by age P and Q is calculated as follows:

$$KLD = \sum_{a=0}^{\Omega} \log\left(\frac{P(a)}{Q(a)}\right) P(a), \tag{1.2.1}$$

with a the age and Ω the maximum age. To get an index summarizing departmental dissimilarities, I calculate the national KLD:

$$KLD_{Nat} = \sum_{i=1}^{N}\sum_{a=0}^{\Omega} \log\left(\frac{P_i(a)}{P_{Nat}(a)}\right) P_i(a), \tag{1.2.2}$$

with i the *département* and N the total of *départements*.

The national KLD is an aggregate indicator for all distributions. I also calculate distortion indices (ID) which highlight the departmental distortions according to age structure. Thus:

$$ID_i(a) = \frac{\frac{P_i(a)}{P_i}}{\frac{P_{Nat}(a)}{P_{Nat}}}. \tag{1.2.3}$$

1.2.2 Aggregation of Data Sources

This study combines two specific sources of data: the first is raw data collected in 19th century censuses. The second is Bonnet (2018b)'s departmental database.

For the period 1851–1900, populations by age group were recorded every 5 years.[1] These data were collected by a team of Franco-American researchers from the University of Ann Arbor and formatted by INSEE.[2] They are available by quinquennial age group, sex and *département*. These data are not as reliable as the 1901–2014 ones, because of the quality of censuses in the 19th century. Bonneuil (1997) explained this point in his study on the demographic transition in French *départements*. Most of the biases come from respondents' poor specification of age (attractiveness for round ages), lack of internal consistency in tabulations (the sum of *départements* is sometimes not equal to the national figure) and bad transcription of the data in tabulations. The impact of these biases is limited because the study focuses on populations by broad age group, and not by single age. Nevertheless, it is important to bear in mind that only the major trends are totally reliable for this period. In order to get populations on January 1^{st} of each year, I assume that the population at the date of the census is similar to the population on January 1^{st} of the census year, and that populations by age group during intercensal periods can be interpolated linearly.

For the period 1901–2014, data come from Bonnet's database (2018b). This companion paper explains in detail the methodology used to estimate age populations. It relies mainly on the protocol of the Human Mortality Database developed by Wilmoth et al. (2007). The raw data that feed the estimation process come for the most part from the archives of the French statistical institutes. They consist of censuses, vital statistics (births and civilian deaths by age) as well as statistics on military deaths and deportations during the two World Wars. In this paper, I use deaths and populations by age as well as births, for each *département* and year.

1.2.3 Unification of the Geographical Framework

French metropolitan borders have little changed over the period 1851–2014. The variations are due to changes in the eastern borders, but also to the integration of new territories. I apply the departmental classification in force from 1918 to 1967, which includes 90 *départements* (see map in Appendix 1.5.1), in order to get a unified geographical framework and compute the inequality indicators. I rebuild the missing *départements* data in this classification. The methods used are different, depending on the database.

Three main territorial modifications took place during the period 1851–1900. *Savoie* and *Nice*'s *comté* integrated France in 1860[3], which resulted in the creation of *Savoie, Haute-Savoie* and *Alpes-Maritimes*. *Alsace-Moselle* integrated Germany in 1870; this led to the creation of *Territoire de Belfort* and *Meurthe-et-Moselle,* that remained under French administration, while *Moselle, Bas-Rhin* and *Haut-Rhin* passed under German administration.[4] Table 1.1 presents these departmental issues and the periods concerned. To rebuild these data, I assume that population changes had been synchronized between the missing *département* and a geographically close *département*. I therefore associate with each missing *département* a reference *département*. The latter is used as a support to estimate missing data. Table 1.1 reveal that these periods are short and the impacts on the overall

[1] Except for 1871 when the census was conducted in 1872 because of the war against Prussia.
[2] Data uploaded on the INSEE website on February 5th, 2018.
[3] This follows the plebiscite of April 22 and 23
[4] Moreover, the former *cantons* of Schirmeck and Saales in *Vosges* joined *Bas-Rhin* in 1870.

results are therefore limited. *Alsace-Moselle* is somewhat different as the missing period is longer. Since reliable data are available before and after, I keep these *départements* in the study.

Table 1.1: Départements with missing data: 1851-1900

Département	Period with missing data
Alpes-Maritimes; Var	1851–1856
Savoie; Haute-Savoie	1851–1856
Vosges; Territoire de Belfort	1851–1866
Meurthe-et-Moselle; Moselle	1851–1866
Moselle; Bas-Rhin; Haut-Rhin	1866–1900

Notes: Periods are the first and the last censuses with missing data. As an example, *Var*'s populations in 1851 and 1856 are not available.

The period 1901–2014 presents two kinds of missing data. The first concerns *Bas-Rhin, Haut-Rhin* and *Moselle*. These *départements* were reintegrated in 1921 in Bonnet (2018b)'s database. I therefore estimate these data during the period 1901–1920. To do so, I proceed as before: I associate a reference *département* (namely *Meurthe-et-Moselle*) to each of these missing *départements* , then I estimate yearly deaths and population by age by assuming that their evolution were synchronized. The second concerns the reorganization of *Ile-de-France* in 1968. Until that date, *Ile-de-France* contained three *départements*: *Seine*, *Seine-et-Oise* and *Seine-et-Marne*. *Seine-et-Marne* remained the same, but the other two turned into seven *départements* (*Essonne, Hauts-de-Seine, Seine-Saint-Denis, Val-de-Marne, Val d'Oise, Paris* and *Yvelines*). I rebuild *Seine* and *Seine-et-Oise* from these new geographical units, by dividing yearly deaths and population by age as well as births pro-rata their distribution in 1968.[5] The sum in the old classification thus remains equal to the sum in the new classification. The relative positions are fixed at their 1968 level.

1.3 Results

1.3.1 Evolutions of Departmental Densities of Population since 1851

I start with the evolution of the French population between 1851 and 2014. Table 1.2 shows the population by sex for several dates. One can see that this evolution has not been linear: two phases can be identified to describe it. I choose the year 1946 as it seems to be the turning point in natural movement trend. During the period 1851–1946, the total population increased by only 10% (Line 3), and only by 5% for men (Line 2). Two causes can explain this evolution. Due to a very early demographic transition, the birth rate in France in the beginning of the 20th century already reached low levels compared to its European neighbors. Coale (2017, p.38) showed that the fertility index dropped sharply since 1820 while this phenomenon appeared rather around 1900 in other developed countries (1890 in Germany, 1913 in Italy). In addition, the mortality increased dramatically during the three major conflicts: the two world wars, and to a lesser extent the war against Prussia in the

[5] Data in both classifications are available for this year.

early 1870s. In addition, these wars impacted strongly the sex ratio: in 1946, the male population is 10% lower than the female population; this difference was only 1 % in 1851 (Line 4). The period 1946–2014 is radically different. In 70 years, the population increased by 60%: the annual growth rate was 65 times greater than the period 1851–1946 one. The baby boom of the post-war years (according to Bonnet (2018b)'s database, the crude birth rate increased from 66 to 90 births per thousand of women between 1936 and 1946) created much larger cohorts than cohorts born before 1946. Moreover, the sharp rise in life expectancy during the second half of the 20th century (female life expectancy at birth increased from 65 years in 1946 to 85 years in 2014 according to the same source) allows older people to live longer, which increases the population. Finally, international migrations (coming mainly from the North-Africa in the 1960s) contribute to increase the population too.

Table 1.2: French population, 1851–2013

	1851	1872	1901	1921	1946	1975	1999	2013
Men	18,112	18,727	19,750	18,455	18,906	25,726	28,443	30,847
Women	18,295	18,894	20,338	20,405	21,014	26,840	30,171	32,852
Total	36,407	37,621	40,088	38,860	39,920	52,566	58,614	63,699
Sex-ratio	99,0%	99,1%	97,1%	90,4%	90,0%	95,8%	94,3%	93,9%

Notes: Numbers in thousand. Sample includes 90 *départements*.

In a second stage I analyze the population density at national level but also for each *département*. The results for several years are presented in Table 1.2. The national density of population in 1851 and 2013 was 67 and 117 inhabitants by km^2, respectively. This increase hides local specificities. The first statement is that the density was multiplied by more than a factor 5 in the most densely populated *département* (namely, *Seine*), while it decreased by 30% in the less densely populated *département* (*Basses-Alpes* in 1851, *Lozère* in 2013). These variations are mainly due to the rural exodus, already highlighted by Ariès (1948). The second is that these evolutions have not been similar for all *départements* during these 150 years. The maximum density increased continuously between 1851 and 2013, but this is not the case for all the others, whose density followed a U-shaped curve: it decreased from 1851 to 1946, then increased from 1946 onwards.

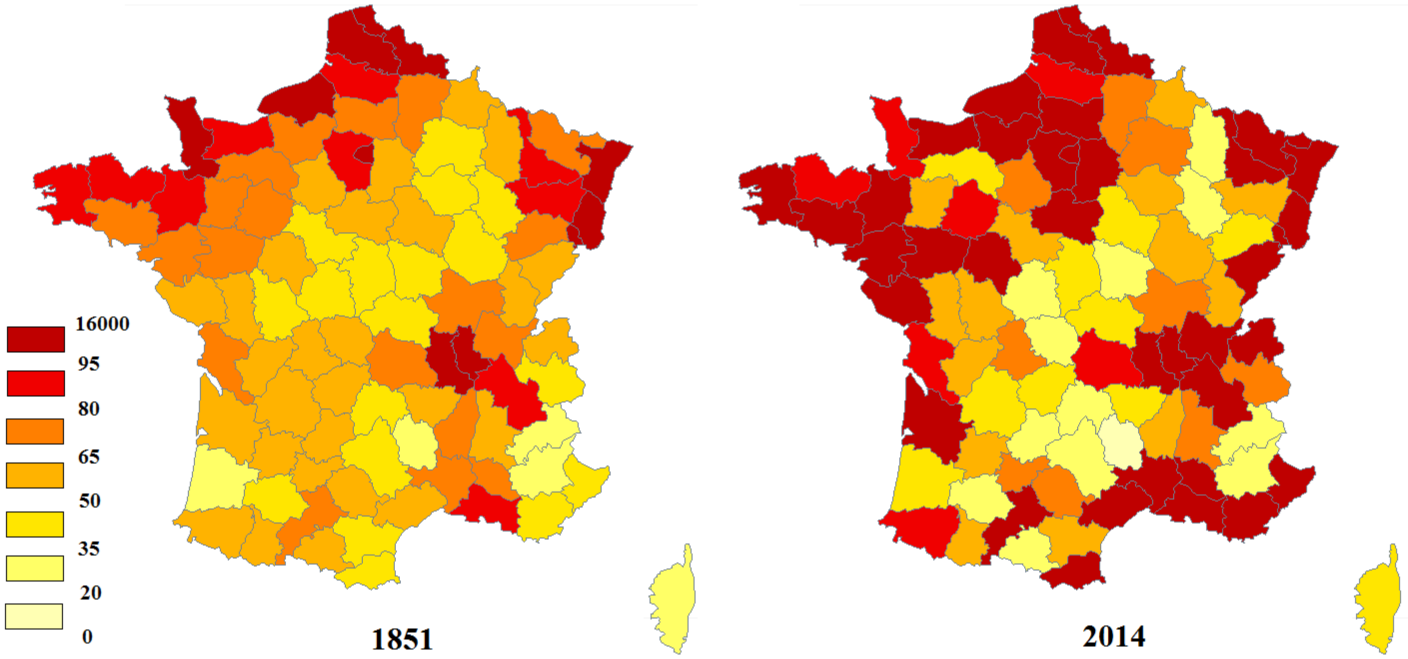
Table 1.3: Population density in french départements

	1851	1872	1901	1921	1946	1975	1999	2013
Min	22	20	17	13	12	15	14	15
25%	49	49	47	42	41	45	48	48
Median	63	60	58	53	53	67	71	75
75%	76	73	75	71	73	104	130	140
Max	2,963	4,625	7,486	9,033	9,579	13,374	14,358	15,655
National	67	69	74	71	73	97	108	117

Notes: Population of both sexes gathered. Sample includes 90 *départements*. 25% means that 25% of *departements* have a density below this level.

In order to better understand the changes of these departmental population densities, Figure 1.3.1 maps the absolute departmental values for 1851 and 2014.[6] Northern France was in 1851 globally more densely populated than the rest of the country. The Channel coasts, the German borders as well as Paris and *Lyon* regions had a population density of more than 80 inhabitants per km^2, whereas these densities were less than 20 inhabitants per km^2 in the Alpine *départements*, *Lozère* and *Landes*. Overall, this finding continues today. One can add the Atlantic and Mediterranean coasts as well as the Swiss border in the densely populated regions, namely where the densities are greater than 95 inhabitants per km^2. The difference between the two maps is mainly in relation to the relative positions: the second revealed a broad band sparsely populated from *Meuse* in the North-East to *Aveyron* in the South-West. This band is called the "empty diagonal" in Oliveau and Doigneau (2016), among others. In 1851, it was not as marked as it is today

Figure 1.3.1: Population density, 1851 and 2014

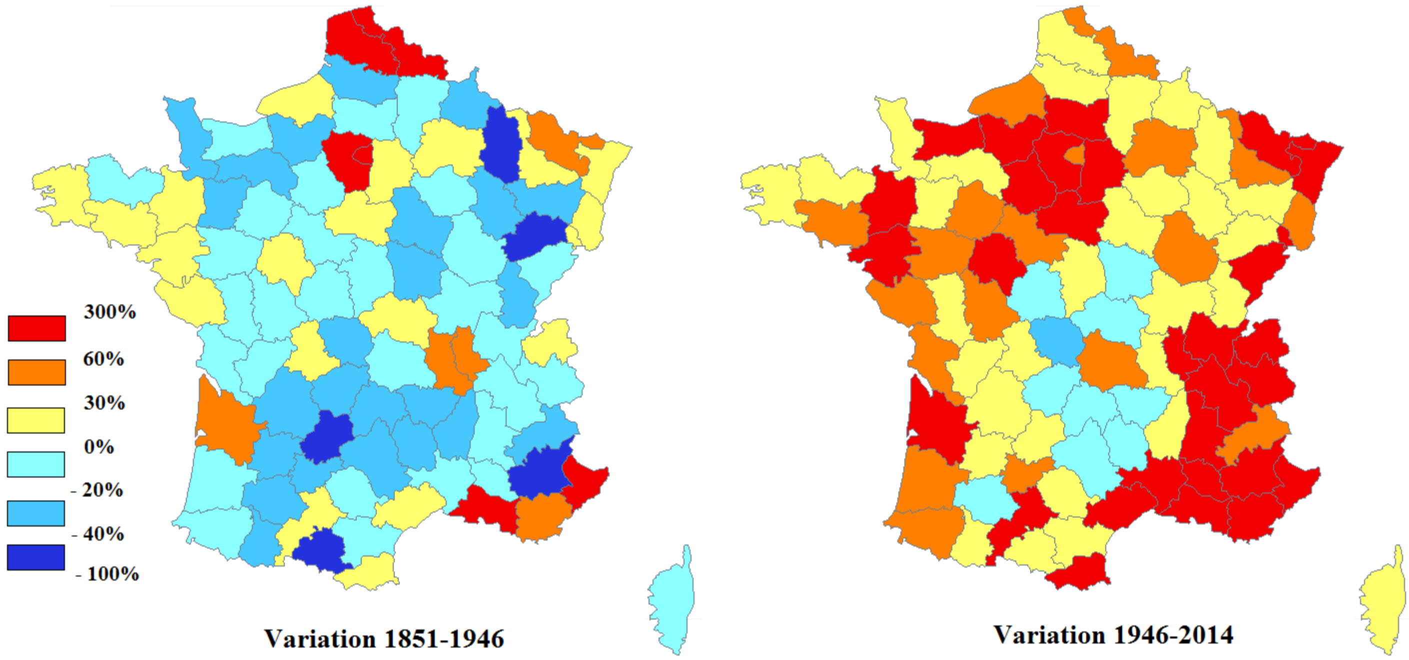


Notes: Population densities based on the population of both sexes gathered. Sample includes 90 *départements*.

Figure 1.3.2 maps the variations of density during the two sub-periods identified from Table 1.2 (namely 1851–1946 and 1946–2014). The previous statements are confirmed. Between 1851 and 1946, the small increase in density was due to four regions: the North of France, Paris and its surroundings, the *départements* of *Loire* and *Rhône* (which host *Lyon* and *Saint-Etienne*), and the Mediterranean coast in which densities have increased by more than 30%. In almost all other *départements*, the population fell or stagnated. The largest decreases are recorded in the Alps or in the North-East (*Meuse* and *Haute-Saône*). From 1946 to 2014, population density fell in a large region around the *Massif Central* while it increased sharply at the national level. The population density in the rest of the country is increasing, with maximum gains reached in a large South-East, in the neighboring *départements* of Paris, and more broadly in the *départements* which host large metropolises (*Haute-Garonne* for *Toulouse*, *Loire-Atlantique* for *Nantes* ... etc).

[6] All the maps were made with the software Philcarto (`http://philcarto.free.fr/`)

Figure 1.3.2: Variation of population density, 1851–1946 and 1946–2014



Notes: Variation of population density based on the population of both sexes gathered. Sample includes 90 *départements*.

1.3.2 The Three Phases in the Evolution of Spatial Distribution of Population

I begin by analyzing the spatial inequality indicators, as presented in Section 1.2.1. Figure 1.3.3 presents the evolution of these indicators since 1851 for the population of women. There was a sharp increase in inequality of population densities since the beginning of the period: the Gini index has more than doubled over the period (from 0.232 in 1851 to 0.320 in 1900 and 0.478 in 2014). While Combes et al. (2011) found a stagnation of the Theil index between 1982 and 2000, the Gini index was still increasing along this period. This can be explained in two ways. Combes et al. (2011) did not weight the *départements* by their area, and the Gini and Theil indices do not give the same importance to each part of the departmental distribution. This illustates why the analysis conducted with my indicators is particularly interesting. Moreover, these indicators highlight three phases in the increase of inequalities for 150 years.

Between 1851 and 1900, the more uneven distribution of the population is explained only by the share increase of the 10% most densely populated km^2: the share of this decile in the total population increased from 21.5% to 30.2%, an increase of about 40%. Inside this decile, the *Seine*'s share[7] – including Paris and its surrondings – doubled (9.1% in 1900) while the one of the remaining decile increased by only 20%. In contrast, the share of all other territories declined, without exception. These evolutions are in line with the rural exodus, whose beginning dated back to the 18th century (Ariès, 1948). Goreux (1956) for example showed how Paris and the other big urban centers attracted agricultural workers during the 19th century; the choice of emigration place was largely explained by the distance. The spatial distribution of the French population can no longer be explained by

[7]The department of *Seine* has an area of 480 km^2, or about 0.01% of the national area

"first-nature" advantages as in pre-industrial societies[8]: agglomeration effects became the main force explaining the changes at work.[9] Thereby, the development of railway during the second half of the 19th century could have favored this process. Mojica and Marti-Hennberg (2011) revealed that in 1880, 90% of agglomerations were connected to the railways. For the authors, the train facilitated migrations of rural people to cities and therefore spatial concentration. Moreover, Fletcher (1961) and Schwartz et al. (2011) showed how lower transport costs at the national level led to the importation of cheaper US wheats. Consequently, the French and the English agricultural sector were plunged into crises between 1870 and 1900. Figure 1.3.3 shows that the share of the four least densely populated deciles strongly decreased from 1870 too: for example, the share of the 20% less densely populated km^2 decreased by only 2% between 1850 and 1870, compared to 11% between 1870 and 1900 (from 11% to 9.7%). I name this phase “hyper-centralization” since the 10% most densely populated km^2 expanded to the detriment of all the others.

Between 1900 and 1968, the 10% most densely populated km^2 were still concentrating the population, even if the World War Two was a temporary break in this process. This break can be fully explained: Bonnet (2018a) showed that internal migrations during this conflict were strong, especially from the North of the country (occupied by the Germans) to the South (in the free zone until 1942). According to this study, the scars left by the conflict in the country’s demography were deep: the refugees who fled densely populated regions such as Northern and Eastern borders, *Bretagne*, *Normandie* and *Seine-et-Oise*, did not fully come back. Nevertheless, over the whole period 1900–1968, the share of the total population who lived in this first decile went from 30.2% to 40%; the cumulative increase reached 80% since 1851 . Unlike the previous period, this hyper-centralization was no longer at the expense of all other territories: the share of the second decile stagnated or increased slightly between these two dates, at around 12%. The population of the *départements* comprising second-tier cities increased as quickly as the national population. On the other hand, the share of the less populated *départements* were still declining. Since 1851, this decrease fell between -20 and -50% according to the deciles. Thus, while the rural exodus affected all territories except the most densely populated between 1851 and 1900, the decline was over in fairly densely populated *départements*. I call this second phase “hyper-centralization thwarted”.

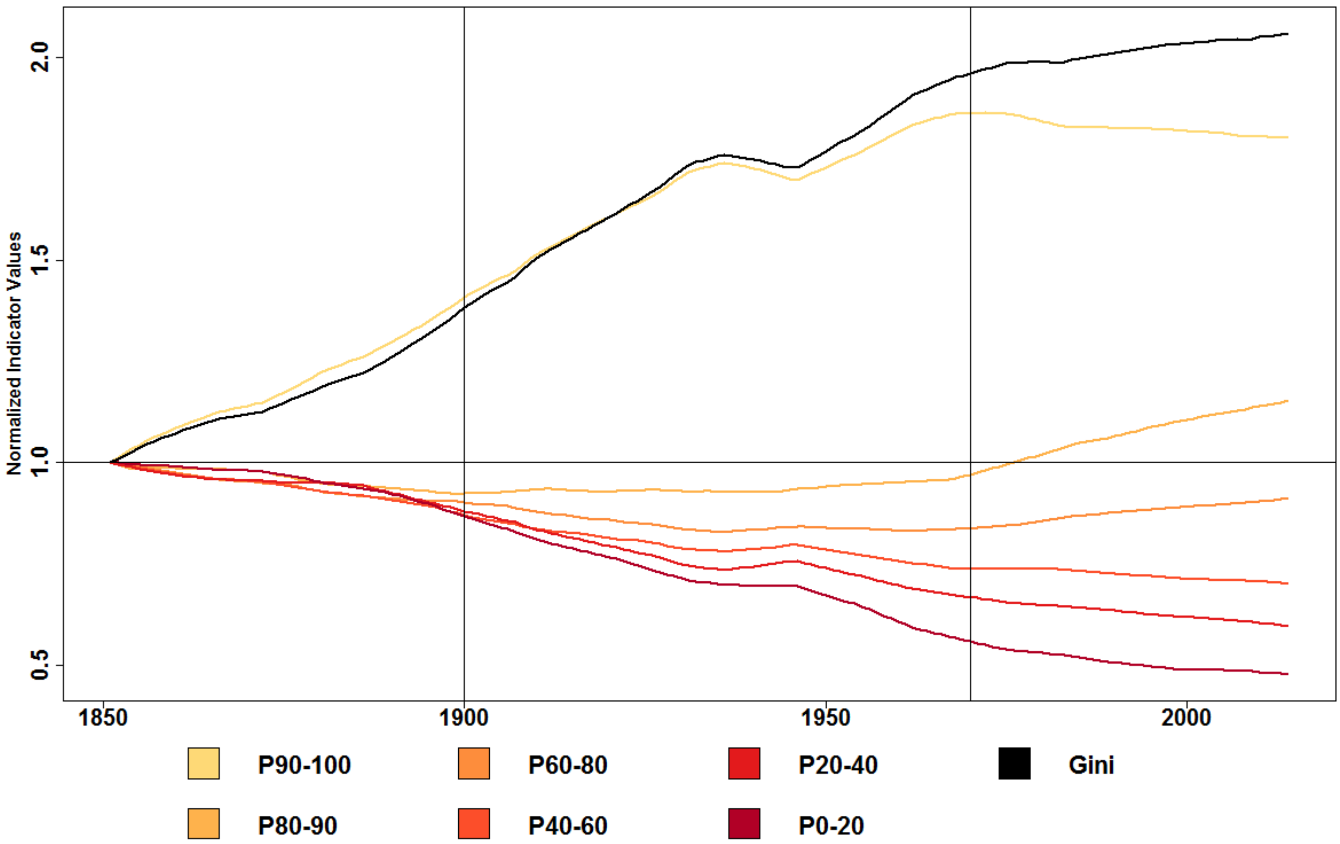
Finally, between 1968 and 2014, the share of the first decile in the total population decreased from 40% to 38.8%, around 3%. The decline is a little more sharp for *Seine*, around 4%. While the share of this *département* in the total population had increased three-fold in a little more than 100 years, the cumulative process stopped for the capital and its suburbs. The second-nature advantage (Krugman, 1993) was no longer enough to attract the national population. This can be explained by congestion costs in the Paris region, in line with the results of Puga (1999), Graham (2007) or Combes et al. (2012). Conversely, the share of the second, third and fourth deciles increased quite strongly. The share of the population living in the second decile went up from 12.3% in 1968 to 14.8% in 2014, an increase of 20%. The share of lower deciles continued to decrease, even if the pace was less sustained. Overall, the rise in global inequalities hides two phenomena: a decrease in the share of the most densely populated *départements* (pushing down inequalities), and a decrease in the share of the least

[8]See Beeson et al. (2001) in the case of US counties

[9]These results are in line with those of Michaels et al. (2017) for USA

densely populated *départements* (pushing up inequalities). I name this phase "multipolarization" of the French population. The multipolarization is geographical and not statistical: *départements* belonging to the second, third and fourth deciles increased over time. One can find in this category *départements* such as *Haute-Garonne, Loire-Inférieure, Haute-Garonne, Haute-Savoie, Isère* or *Gironde*, scattered throughout the landscape. It is within these *départements* that the second-tier cities such as *Toulouse, Bordeaux, Nantes* and *Grenoble* are located. The expression "Paris and the French desert" developed by Gravier (1947) gradually loses its importance. The policy pursued by DATAR since the 1960s succeeded. It allowed the displacement of the most mobile jobs towards regions where mass unemployment threatened at the end of the "30 glorious".

Figure 1.3.3: Spatial inequalities of population density, 1851–2014



Notes: Computations based on the population of women. P90-100 refers to the share of national population who lived in the 10% of km^2 with the highest density values. All inequality indicators are weighted by km^2 and normalized by 1851 values. Sample includes 90 *départements*.

The share of the less densely populated *départements* declined since 1851, with a temporary improvement due to World War Two. This evolution, which is explained by a population growth rate below the national average, can come from two factors. The first is related to natural movement: the growth rate of the population from a surplus of births to deaths is lower than the national rate. The second is related to migratory flows: the growth rate of the population coming from net migrations is lower than the national rate. *Départements* can be classified in four categories as presented in Table 1.4. Class 1 *départements* are those with both rates below the national average. Class 3 *départements* are those with both rates above the national average. Classes 2 and 4 are mixed situations.

Data from the period 1851–1900 do not allow for this classification since they do not include the total births and deaths of each *département*. Conversely, Bonnet (2018b)'s database is sufficient. By difference between the population change between two dates and the natural movement, it is possible to compute the apparent net migratory flow. Figure 1.3.4 presents the classification for periods 1901–1968 and 1968–2014, for the population of women. These two phases correspond to what I have called "hyper-centralization thwarted" and "multipolarization".

Table 1.4: Classification of french départements according to migrations and natural movement

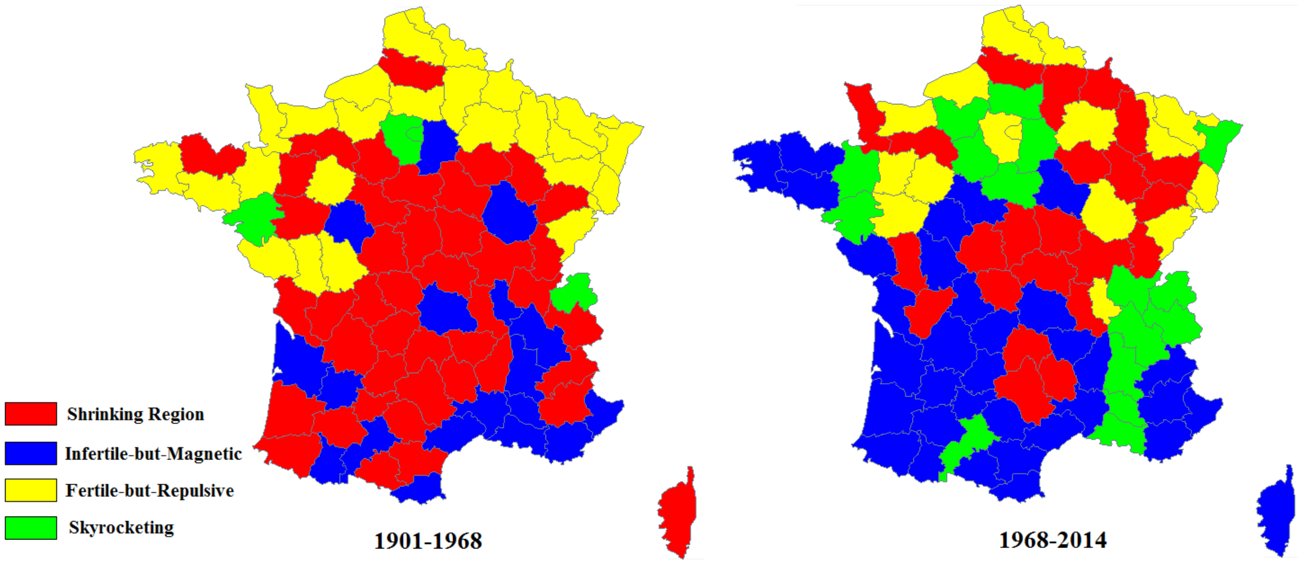
		Net Migration Rate	
		Upper	**Lower**
Intrinsic Growth Rate	**Upper**	Skyrocketing	Fertile-Repulsive
	Lower	Infertile-Magnetic	Shrinking regions

Notes : Classification of the French *départements* according to their population evolution rates due to intrinsic growth or migratory movement, relative to national values.

According to the period 1901–1968, Figure 1.3.4 reveals a deep difference between the North and the South according to the natural movement ("Skyrocketing" and "Fertile-but-Repulsive" *départements* in green and yellow against "Shrinking" and "Infertile-but-Magnetic" *départements* in red and blue). In the North (generally above an arc connecting *Doubs*, *Seine* and *Vendée*), the growth rate of population due to the natural movement is higher than the metropolitan rate, different from the South. This region was named *"croissant fertile"* (see for example Francart, 1983) and was visible since the Second World War. This phenomenon compensated a net migration rate lower than the national average: these regions were not attractive for migrations (whether internal or coming from abroad). In the South, *départements* along the Mediterranean coast, the Rhone Valley and those which host major cities compensated this impairment by a strong attractiveness: net migration rates are higher than the national average. For all the others, in red on the map, the disadvantages cumulated: their shares fell because of a weak natural movement and a lack of attractiveness on the migratory side.

The situation changed during the period 1968–2014. For this period, researchers have studied migration flows, and in particular interregional flows. This work has been carried out on each intercensal period (Baccaïni (2001), Baccaïni and Lévy (2009) for the most recent studies), but also dynamically between 1954 and 2008 (Baccaïni and Dutreuilh, 2007). The first striking result concerns *Seine* and *Seine-et-Oise* and feeds the conclusions already stated above. Migration rates fell below the national ones, which explains the decline in the share of the first decile (Figure 1.3.3). This statement is supported by the fact that the neighboring *départements* became "Skyrocketing" ones: they draw a green belt around *Seine* and *Seine-et-Oise*. These *départements*, more and more efficiently linked by transport to the capital, became attractive for migrations. I name it the "wide belt of attractiveness". While Baccaïni and Dutreuilh (2007) put forward the reversal concerning *Ile-de-France*, the authors could not mention this departmental belt since they worked at the regional level. The second observation is about the North-South gradient concerning natural movement, which is no longer as readible as before. It became stronger than the national average in the South-East. The reverse occurred in *Bretagne* and in *départements* like *Meuse* and *Ardennes*, while they were at the heart of the baby boom following the Second World War. The North-West and the North-East differed on the migratory side: the former was attractive, which was not the case of the latter. These results support the idea that *Aisne*, *Ardennes*, *Meuse* or *Somme* constitute a "shrinking region" in the sense of Oswalt and Rieniets (1984). This region, which had enjoyed a first-nature advantage (Krugman, 1993) for much of the 20th century (mines, heavy iron and steel plants), is still facing difficulties in converting its productive capital. Laménie (2016) showed, for example, that the population of *Ardennes* decreased by 10% between 1968 and 1999, due to the emigration of young people. This process led to a drop in the birth rate and an aging population. Finally, the South-West became attractive for migratory flows, which was not the case previously: *Dordogne*, *Lot-et-Garonne*, *Tarn-et-Garonne* or *Corrèze* are examples of *départements* that switched from “Shrinking Regions” to “Infertile-but-Magnetic” class. Baccaïni and Dutreuilh (2007) noted this attraction for the South and West from the 1960s, which contributed to the spatial redistribution of the population. Nevertheless, one can see on Figure 1.3.4 that some *départements* remained on the margins of this process. This is the case in the south of *Massif Central* (*Aveyron*, *Cantal*, *Lozère*), but also in *Deux-Sèvres* and *Charente* which did not benefit from the amenities of Atlantic coast.

Figure 1.3.4: Classification of french départements, 1901–1968 and 1968–2014



Notes: Computations based on the population of women. Classification built according to the theoretical classification presented on Figure 1.3.4. Sample includes 90 *départements*.

1.3.3 The Uneven Spatial Distribution of Population According to Age

The long-term studies conducted so far could not analyze differentiated trends in the spatial distribution of the population by age group. The data available in this study allow this analysis. Overall, I find that the trends of population aged 0 to 19, 40 to 49 and 50 to 64 are the same as the trends observed for the whole population. On the other hand, they are significantly different for women aged 20 to 29, 30 to 39, 65 to 79, and 80 and over. Figure 1.3.5 reveals the inequality indicators for women aged 30 to 39 and 65 to 79.

According to women aged 30 to 39, the main difference with national trends comes from the period 1990–2014, during which the share of the first decile increased while the share of the third and fourth deciles stagnated. The evolution of the first decile was almost completely due to the increase of *Seine*'s share (Paris and surrondings). Consequently, France is facing a new phase of "hyper-centralization thwarted" concerning this age group. It gathers individuals at the heart of their working lives, usually with high salaries and stable work situations. This result supports Combes et al. (2011)'s paper, which show that the Gini index of the spatial distribution of tertiary value added followed an inverted U-shape from 1860 to 1982 and increased from 1982 to 2000. The parenthesis in the aggressive spatial planning policy counducted during the 1960s could explain this process. Following works on endogenous growth, economists and politicians were aware of how economically strong regions have to be supported in order to redistribute income to poorer geographic areas (see, for example, Jayet et al (2006) and Davezies (2008)).

For the older ones, changes are different for two reasons. Overall, the Gini index increased only by 60% over the period 1851–2014, compared to 110% for the whole population. From 1851 to 1910, this index remained stable, hiding contrary evolutions: both the share of the first decile and the share of the less densely populated *départements* expand to the detriment of the second, third and fourth deciles. Thus, there was a deconcentration of the elderly population in France from 1851 to 1900, unique in my statistics of population by age. Between 1900 and 1968, the hyper-centralization thwarted was at work: the share of the 80% least densely populated km^2 fell, while the share of the first decile increased and that of the second stagnated. Finally, between 1968 and 2014, the multipolarization appeared: the share of the most densely populated territories in old people decreased, while the one of the *départements* of the second, third and fourth deciles increased.

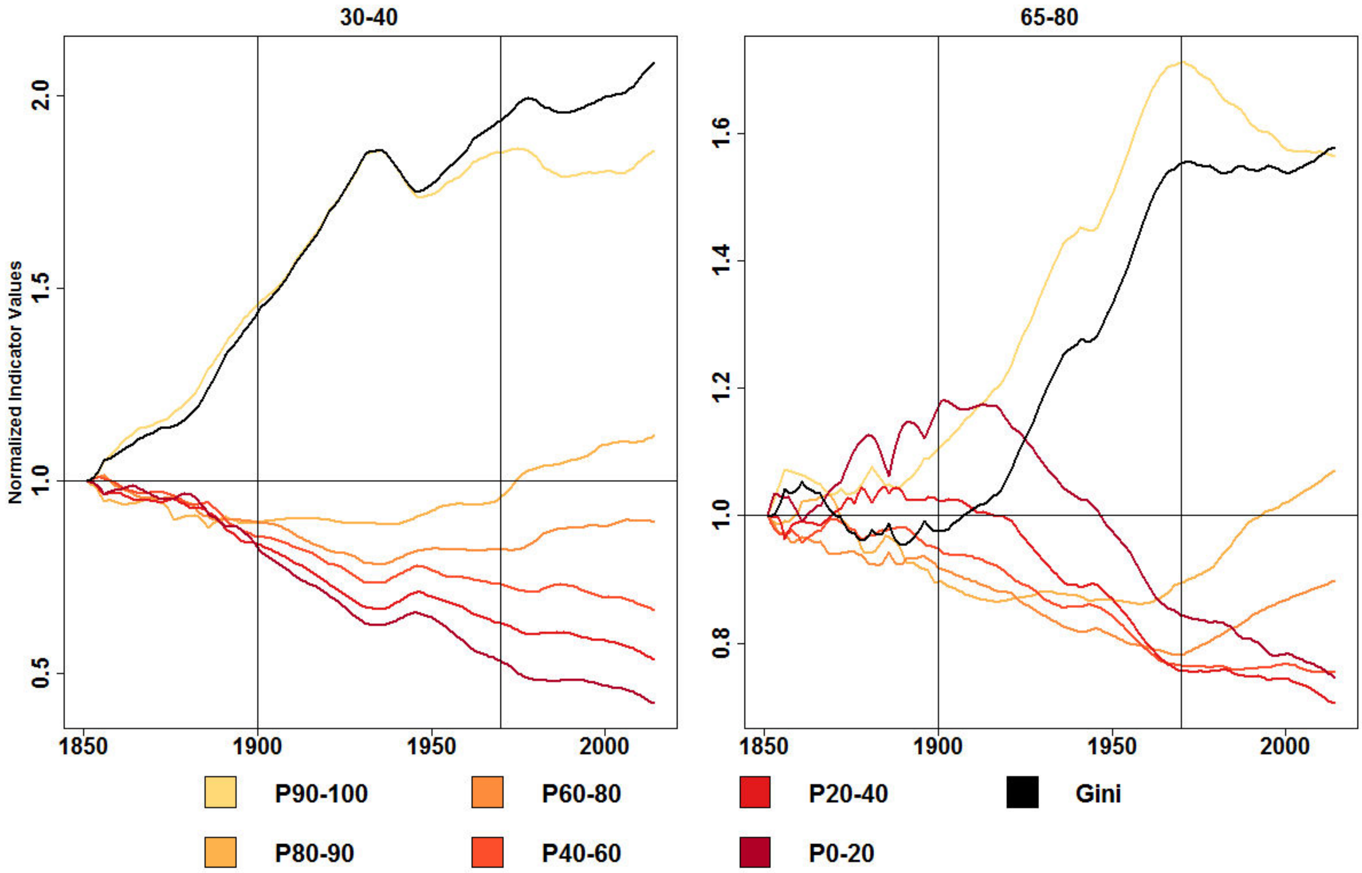
These results are quite the same for men (See Appendix 1.5.2 for the corresponding graphs).

Beyond the evolution of the spatial distribution of population by age, it is interesting to know which are the most unevenly distributed populations, and whether these relative positions have evolved over time. Figure 1.3.6 presents the age profile of the Gini index for women and several dates along the 150 years of this study.

First of all, inequalities have increased for all age groups, which is consistent with what was presented earlier. These increases are significant: if one consider the age group 0–19, the Gini index went from a value of 0.25 in 1851 to a value of 0.48 at the end of the period. With regard to the age profile of the Gini index, Figure 1.3.6 reveals that the population density inequalities in 1851 were similar for all age groups between 0 and 64-year-old, with a Gini value of about 0.25. Beyond these ages, the values were growing. In other words, women aged 65 and over were much more unequally spread than the others. This specificity of the oldest ages gradually disappeared during the end of the 19th century: the inequalities observed for this age group became the weakest from 1901 onwards. Conversely, at this date, the flat profile between age 0 to 65 disappeared too. The profile reveals an inverted U-shape, more and more pronounced over time, where the most uneven age group is 20–29.

For the period 1851–1900, this profile can be explained as follows. Spatial inequalities of population density were the same from 0 to 64-year-old because the working age groups remained in the territory

Figure 1.3.5: Spatial inequalities of population density by age group, 1851–2014

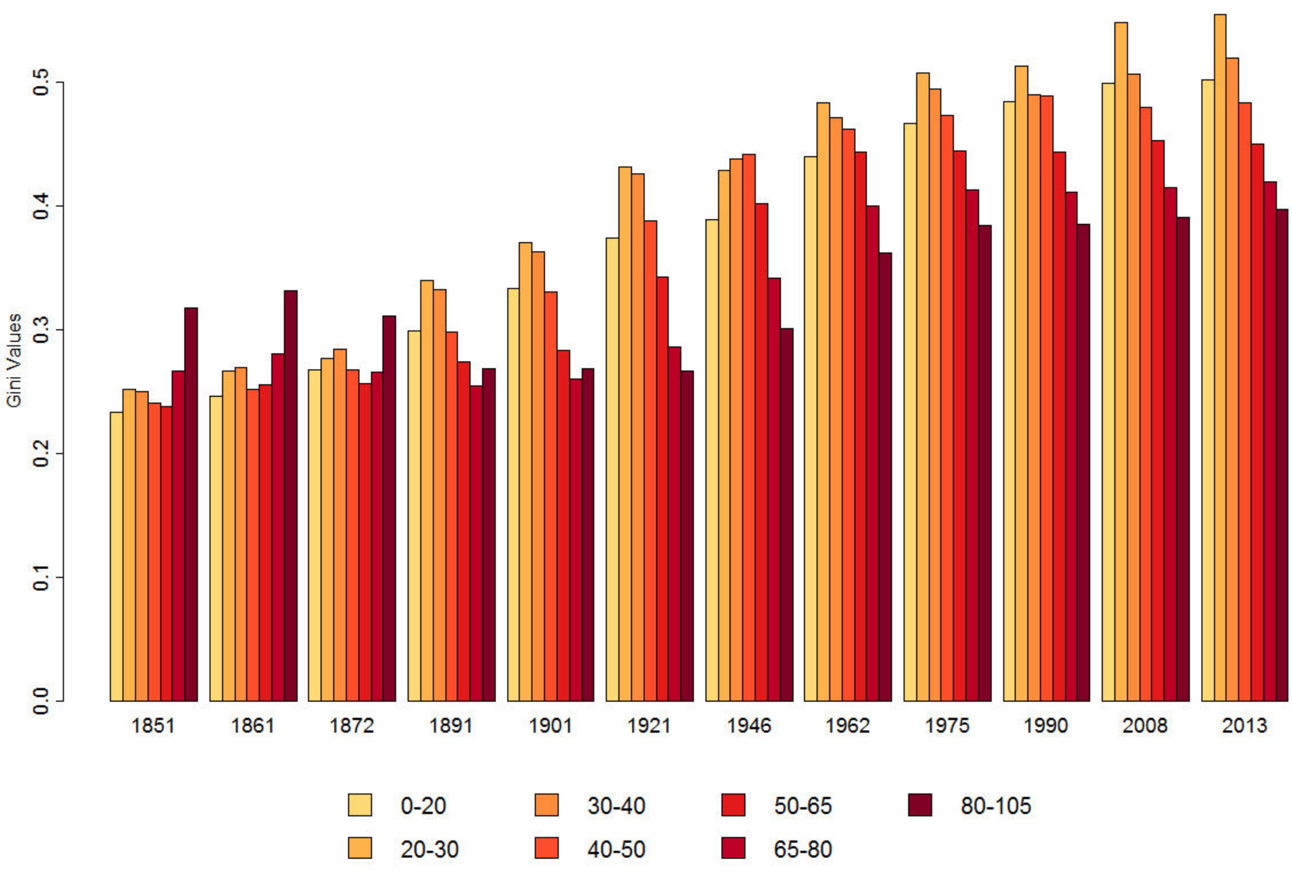


Notes: Computations based on the population of women. P90-100 refers to the share of national population who lived in the 10% of km^2 with the highest density values. All inequality indicators are weighted by km^2 and normalized by 1851 values. "30–40" refers to age 30 to 39 (40 excluded). Sample includes 90 *départements*.

where they were born. Even if this observation is less true concerning the Paris region (the inverted U shape is visible since 1851 for *Seine*), it applies to the French territory as a whole. In this context, the strong inequalities concerning women aged 65 and over did not come from a choice of location but from spatial differences in mortality. In fact, the departmental shares were related to the mortality rates at these ages: there were fewer elderly people where their mortality rates were very high. This is the case for example in *Bretagne* and in the Alps, where life expectancy was very low compared to the national average. For the period 1861–1865, female life expectancy at birth was 22 in *Basses-Alpes* and 28 in *Finistère* according to Bonneuil (1997). Conversely, they were overrepresented in a broad band linking *Normandie* and *Meuse*: in *Yonne, Aube, Meuse* and *Eure*, life expectancy at birth for the same period was 55, 53.5, 52 and 50 years, respectively. Subsequently, this phenomenon progressively disappeared: spatial mortality inequalities decreased strongly from 1881 according to Bonnet and d'Albis (2018). They name this phenomenon "Centennial Convergence". These evolutions explain the specificity of Figure 1.3.5: the share of the least densely populated territories in elderly people increased initially because of the survival of a larger number of individuals at these ages. So there was no rural counter-exodus. For these reasons, I name that differentiated spatial concentration of elderly people "undergone concentration".

The inverted U-shaped profile that appeared in 1901 can be explained differently. From this date, the rural exodus was powerful: the 20 to 39-year-olds moved to the most densely populated areas to find a job, leaving rural areas that no longer had to offer them. This migration did not exist among older people who remained in their home territories, slowing down the unequal process for these age groups. On the opposite of the "undergone concentration" experienced by the elderly, this inverted U-shape profile results from a chosen process of territorial polarization. 20 to 39-year-old were strongly attracted by the capital and then by second-tier cities during the period 1901–2014. This was less the case for other age groups, whose location choices were less constrained. This is particularly true among older people today: their income stream is not conditioned by their location, as it is the case for the youngers. They can settle everywhere in the territory, and especially in the rural *départements*, less densely populated and with attractive amenities. This age profile is accentuated nowadays due to a sharp rise in inequality among 20 to 29-year-old. This age group is today the most unevenly distributed. As such, the 10% most densely populated km^2 host 43% of women aged 20 to 29. For those aged 65 and over, this figure is only 32%.

Figure 1.3.6: Age profile of gini index for population density, 1851–2014

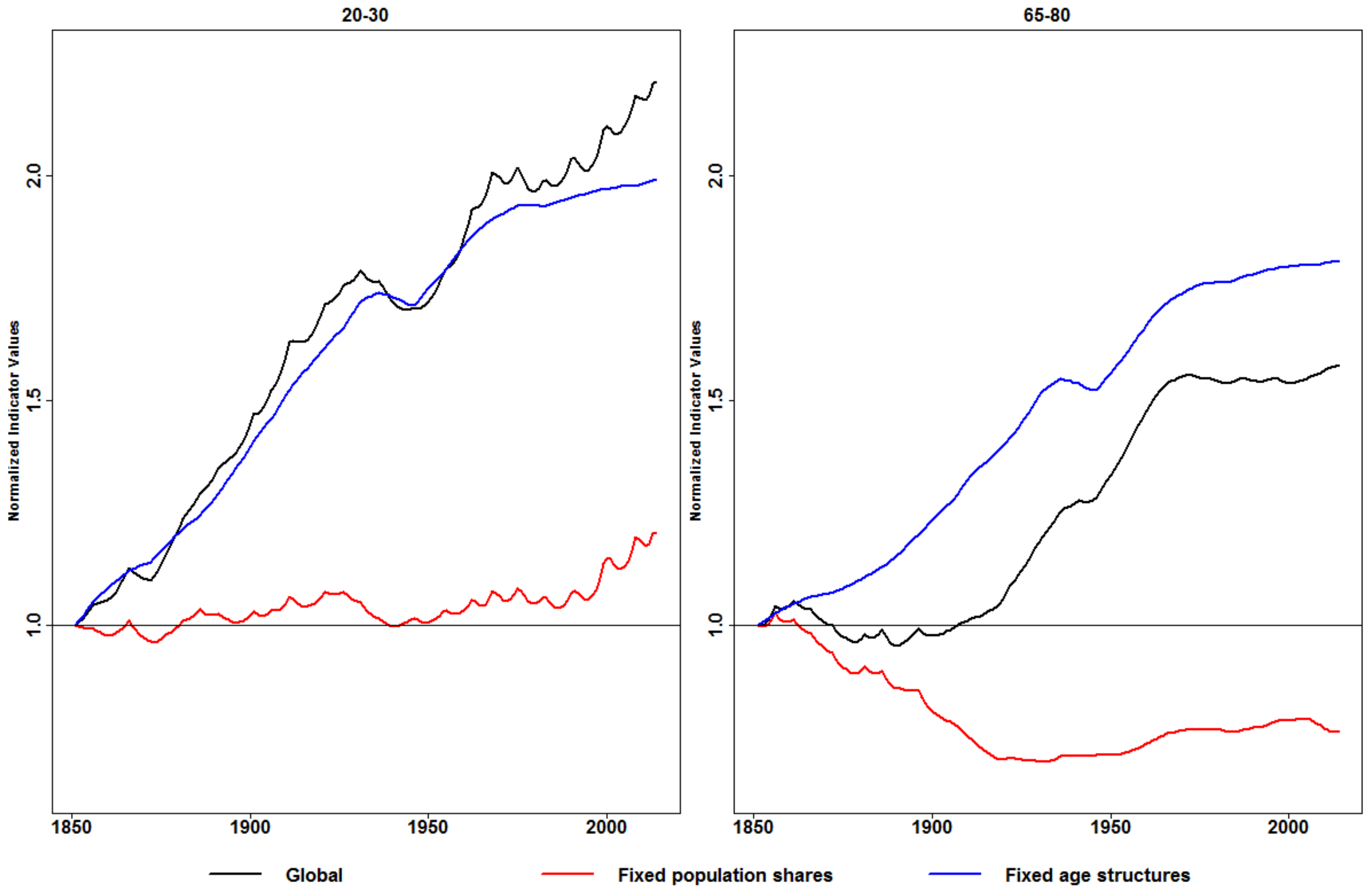


Notes: Computations based on the population of women. "30-40" means age 30 to 39 (40 excluded). Sample includes 90 *départements*.

By comparing Figures 1.3.5 and 1.3.6, one can see that the spatial distribution of the population according to the age groups followed both a common and a specific process. For example, the population of the elderly is more unequally distributed in 2013, like all the other age groups, but the evolution was different than the evolution of the youngest. These differences can be explained by changes in age structures: if the age structures remained the same, then the evolution of spatial inequalities would be the same for all age groups. On the other hand, if the age structures were more and more differentiated between the *départements*, then the changes would be different according to the age groups. For example, if the share of younger people decreased in the *départements* where the share of the population in the national total decreases, then the increase in spatial inequalities would be greater than the one observed for the whole population.

To show that departmental age structures matter in the evolution of spatial inequalities, I dissociate two different effects (See Appendix 1.5.2 for the formal explanations of the method used). The first is the background process induced by the evolution of total population densities. The second is the specific process explained by the evolution of population age structures in *départements*. To do so, I fix in stage 1 departmental populations at their initial level and allow age structures to vary. Then, in stage 2, I fix age structures at their initial level and allow departmental populations to vary.

Figure 1.3.7: Spatial inequalities of population density by age group and type, 1851–2014



Notes: Computations based on the population of women. "Fixed population shares" means that the shares of *départements* are fixed at their 1851 levels. "Fixed age structure" means that the age structure of each *département* is fixed over the 1851–2014 period. "20-30" means age 20 to 29 (30 excluded). Sample includes 90 *départements*.

Figure 1.3.7 presents the evolution of the Gini index for women aged 20 to 29 and 65 to 79. For the first age group, the red curve shows that the change in departmental population age structures contributed to the rise in spatial inequalities. They would have increased by 20% if the total departmental populations had remained stable. In other words, the 20 to 29-year-old shares are more strongly differentiated than in 1851, and this differentiation has accelerated since 1990. On the other hand, the change in age structures of the 65 to 79-year-old has slowed the rise in inequality. The curve in red has a U-shape: until the 1930s, the 65 to 79-year-old shares homogenized; they differentiated from 1930 onwards. More broadly, this profile is observed for all women aged 65 and over.

1.3.4 Differences in Departmental Structures of Population by Age

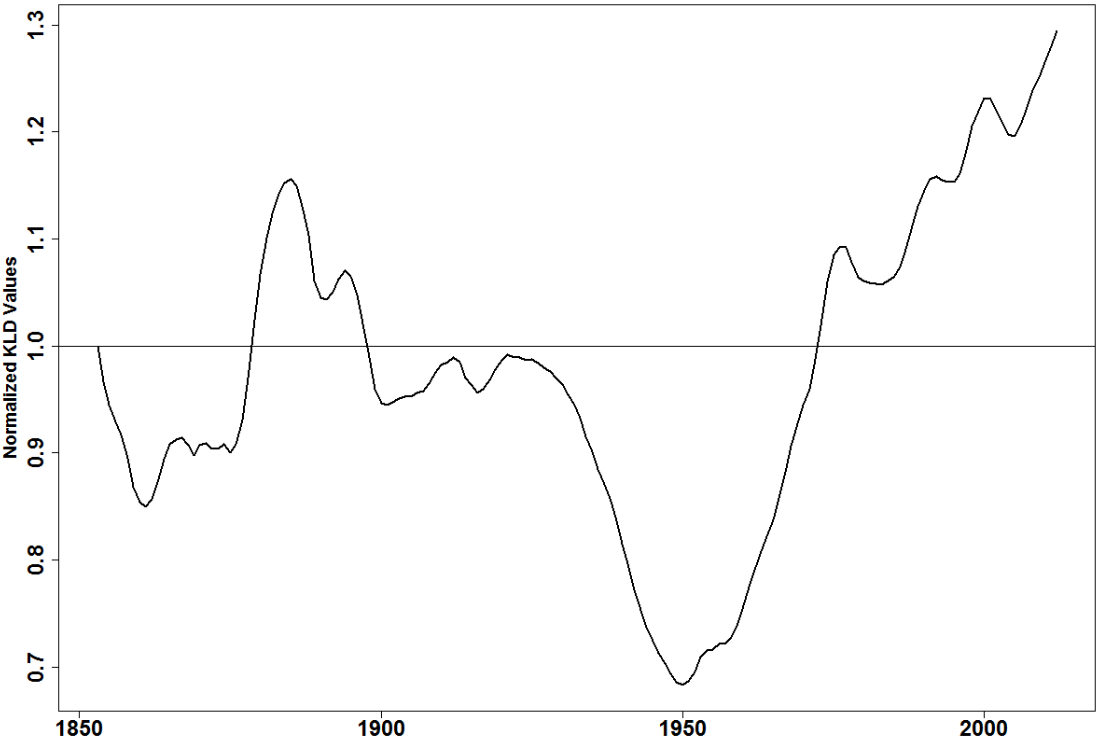
National age structures of population have changed since 1851: at that time, the women aged 65 and over accounted for only 7% of the female population, compared with 21% today. For those under 20, these shares are 35% and 23%, respectively. These results are different at the departmental level. Figure 1.3.7 shows that these differentiated evolutions have impacted the spatial distribution of the population for each age group.

At first I am interested in the overall dissimilarity of population age structures. I use the Kullback-Leibler divergence calculated at the national level (equation 1.2.2), namely KLD_{Nat}. The inequality indicators presented until now weighted the *départements* by their area. Here, the weight assigned to each *département* is the same. Figure 1.3.8 presents the evolution of KLD_{Nat}, for the population of women. Its evolution has not been monotonous over the period. To describe it, I define three sub-periods: 1850–1930, where the index remained broadly stable, 1930–1950 where it declined sharply, and 1950–2014 when it increased by 85% and reached a 30% level higher than the 1850s one. In order to provide an explanation for these contrary variations over the past 80 years, I use the additivity property of KLD_{Nat}. One can break down this indicator in order to get the contribution of each *département* in its variation between two years. This was done for the two sub-periods 1930–1950 and 1950–2014 identified in Figure 1.3.8. I map the contributions of each *département* in Appendix 1.5.4.

Between 1930 and 1950, France experienced a territorial homogenization. Going into more detail, the departmental age structures homogenized due to four distinct geographical areas (in red on Figure 1.5.2): *Seine*, *Bretagne*, *Yonne* and its neighboring *départements*, the South-West along the *Garonne* valley. To better understand what may have caused this homogenization, I represent in Figure 1.3.9 the distortion indices according to four *départements,* which represent these geographical areas. In order to avoid the issues of small numbers in the computations of departmental distortion indices, "65–79" and "80 and over" age groups have been gathered. Moreover, these distortion indices have been smoothed over 5 years.

Figure 1.3.9 reveals that the decline in dissimilarity between 1930 and 1950 is caused by two main factors. The first is the decrease of the overrepresentation of the oldest in Center-East and South-West (*Yonne* and *Tarn-et-Garonne*), coupled with the increase in the share of these age groups in *Bretagne* (*Finistère* for example). This phenomenon is partly explained by the process of convergence of health conditions, very powerful during the inter-war period *(Bretagne* quickly caught up, as shown by d'Albis and Bonnet, 2018). This process could be explained by migratory flows too. Bonnet (2018a)'s work show that migrations occured from the the whole North to the South and more specifically to the South-West. It is likely that internal migrations during the war mainly concerned active age groups, who were in the capacity to move. Consequently, the overrepresented

Figure 1.3.8: NORMALIZED KLD_{Nat}, 1851–2014

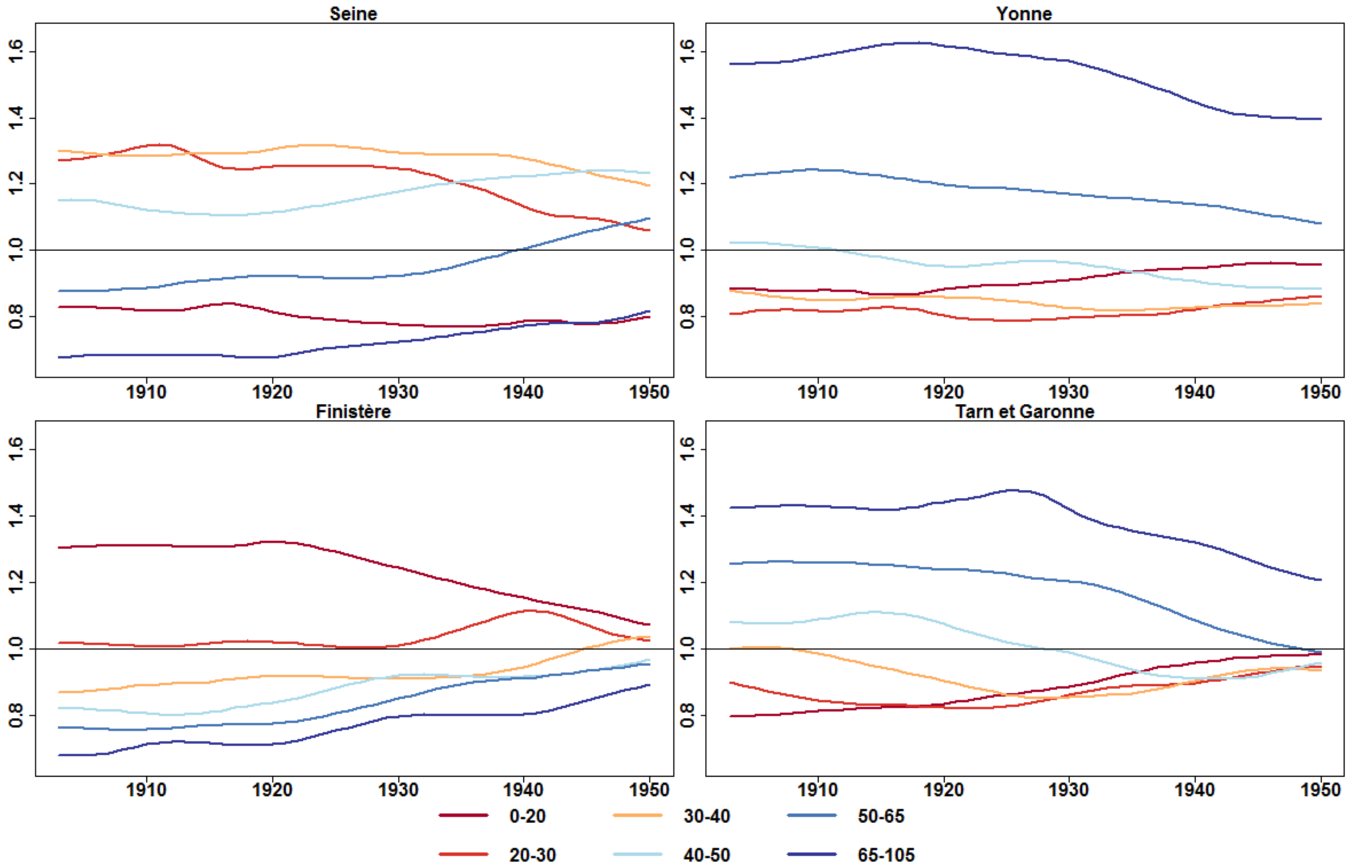


Notes: Computations based on the population of women. KLD_{Nat} is computed according to equation 1.2.2: it is the sum of dissimilarities between *départements* and France as a whole. KLD_{Nat} is non-weighted by population. Sample includes 90 *départements*.

young people in *Bretagne* moved to the South-West where they were underrepresented. The second reason for the decline in overall dissimilarity is specific to *Seine*: the distorsion index according to women aged 20 to 29 fell sharply, losing 20 points over the period. This decline began before the Second World War, which suggests that the conflict is not directly the cause.

Conversely, since 1950, France faced a spatial specialization process which is still on the rise. This sharp increase came from four distinct geographical areas, namely *Seine-et-Oise*, *Hautes-Pyrénées*, *Nièvre* and a large region around *Corrèze* in the South-West. Figure 1.3.10 represents the distortion indices for these four *départements*. One can see that the age structures of the three rural *départements* (*Nièvre, Hautes-Pyrénées, Corrèze*) have diverged very strongly from the national average due to two parallel phenomena: the increasing overrepresentation of the elderly, and the growing underrepresentation of young workers and their children (0–39 years). In *Nièvre*, the share of women aged 65 and over was 40% higher than the same share at the national level. In 2014, this share was 29.3% compared to 20.5%. At the same date, the share of women aged 20 to 29 is 30% lower than the share at the national level, whereas it was only 4% in 1950. The situation is opposite in *Seine-et-Oise*: the distortion indices, which are quite low at the beginning of the period (between 90 and 110%), diverge continuously afterwards. The share of the elderly in the total population is nearly 40% lower than the national share, while the share of 20 to 39-year-old is 20% higher. According to these results, France is differentiating more and more: the country opposes a rural France in which pensioners are overrepresented, and an urban France where the active age groups are overrepresented. This is again in line with the new geographical economy, which postulates an agglomeration of service jobs

Figure 1.3.9: Distorsion indices, 1901–1950



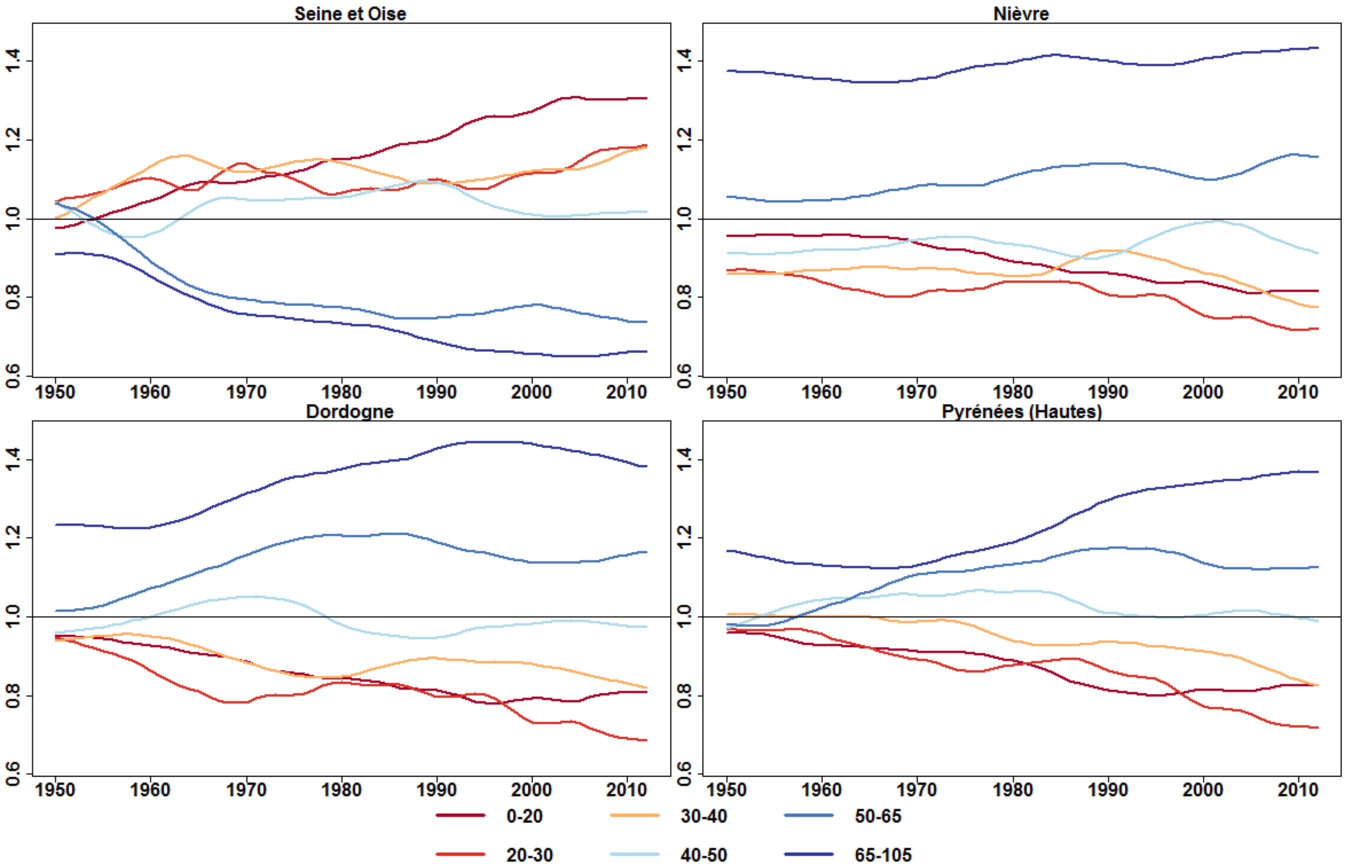
Notes: Computations based on the population of women. Ditorsion index for age x is the ratio between the population's share of age x in one *département* and the same share at the national level. These distorsion indices are smoothed over 5 years. "30-40" means age 30 to 39 (40 excluded). Sample includes 90 *départements*.

in major cities, in parallel with an increase in wages and real estate prices. These developments are gradually chasing retirees whose income is by definition lower. The elderly settle in less densely populated areas such as the South-West, the Center or the Atlantic coast, creating in these territories a residential economy strongly focused on personal services or tourism (Davezies, 2008). As such, Baccaïni and Dutreuilh (2007) showed that internal migrations were positive at all ages except 20 to 29 in *Bretagne*, *Pays de la Loire*, *Limousin*, *Aquitaine* and *Poitou-Charentes* between 1999 and 2004. On the other hand, these internal migrations were negative for all ages except the 20 to 29 age group in *Ile-de-France*.

Finally, I conduct a cartographic analysis of the territorial specialization. Figures 1.3.11 and 1.3.12 show the departmental distortion indices for the female population in 1946 and in 2011 for women aged 20 to 29 and those over 65. These years are respectively the lowest point and the highest point in the process of territorial specialization. Moreover, these two age groups are the most emblematic of the process. Values of these distortion indices in 1856, 1896, 1946 and 2011 are available in Appendix 1.5.5.

Figure 1.3.11 shows these distortion indices for women aged 20 to 29. In 1946, the situation was very homogeneous: indices are almost all between 0.9 and 1.1. The situation was quite different in 2011. First of all, the population aged 20 to 29 was largely overrepresented in *départements* with

Figure 1.3.10: Distorsion indices, 1950–2014

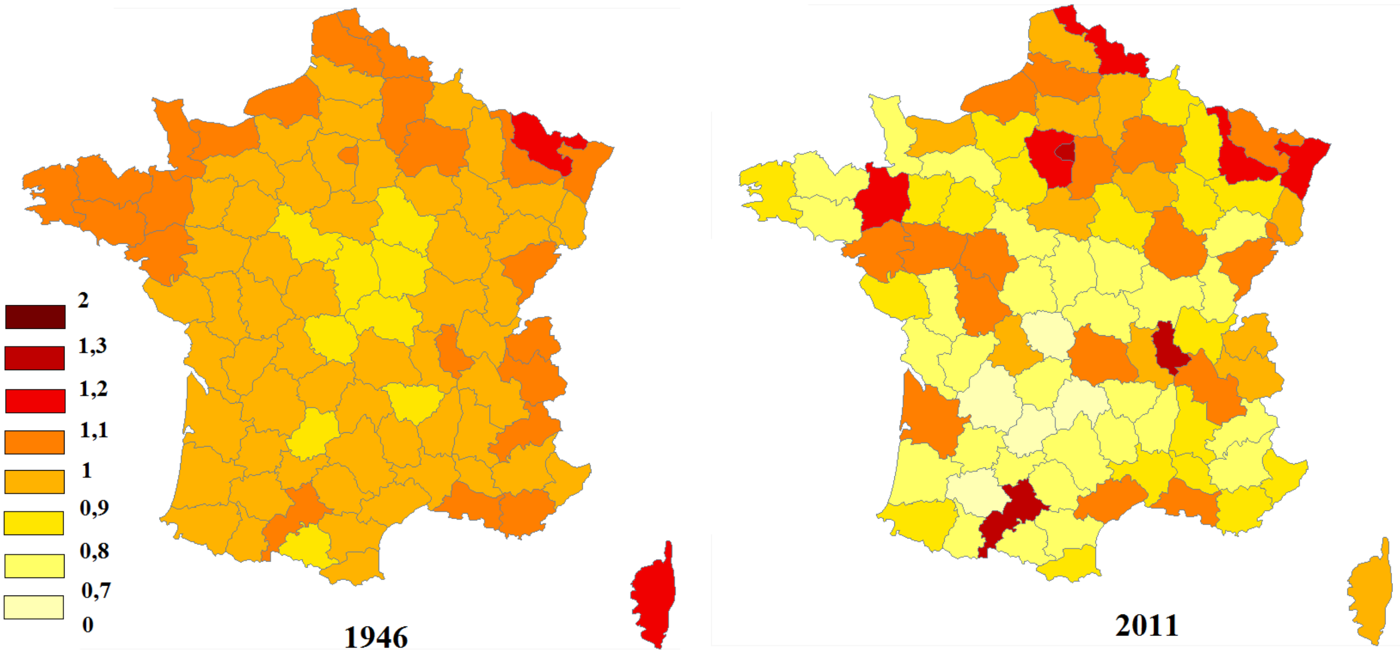


Notes: Computations based on the population of women. Ditorsion index for age x is the ratio between the population's share of age x in one *département* and the same share at the national level. These distorsion indices are smoothed over 5 years. "30-40" means age 30 to 39 (40 excluded). Sample includes 90 *départements*.

big cities: the Paris region already seen in Figure 1.3.10, but also *Haute-Garonne (Toulouse)*, *Rhône (Lyon)* or *Ille-et-Vilaine (Rennes)* in which these indices are comprised between 1.2 and 1.4. Other *départements* appear with less pronounced situations: *Hérault (Montpellier)*, *Gironde (Bordeaux)*, *Côte d'Or (Dijon)* or *Loire-Atlantique (Nantes)*. Conversely, the proportion of young adults in the population was low in central France and in *Bretagne*. The minimum values are reached in South-West *départements*, such as *Cantal* or *Dordogne*. The distortion indices show that this share was globally 35% lower than the national one, and has been steadily declining for 60 years. These *départements* are thus becoming exclusion lands for young adults.

Figure 1.3.12 reveals the distortion indices of women aged 65 and over. The situation in 1946 is different from the one of the youngest since it presents a clear and marked spatial configuration. A large South-West, from *Pyrénées-Atlantiques* to *Haute-Saône* via *Limousin*, showed high distortion indices, with peaks reached in *Creuse*, *Ariège* or *Nièvre* (around 1.4). Conversely, the shares in the whole Northern France as well as Eastern borders are lower than the national average. In 2011, the situation is generally similar to the one which prevailed 60 years ago. This suggests that the geography of old age is deeply rooted in territories. The South-West of the country still retains a population among which the oldest are overrepresented. However, mirroring what is observed for 20-29 year olds, urban *départements* stand out: *Haute-Garonne*, *Gironde* or *Hérault* have lower

Figure 1.3.11: 20–29 Year-old distorsion indices, 1946 and 2011



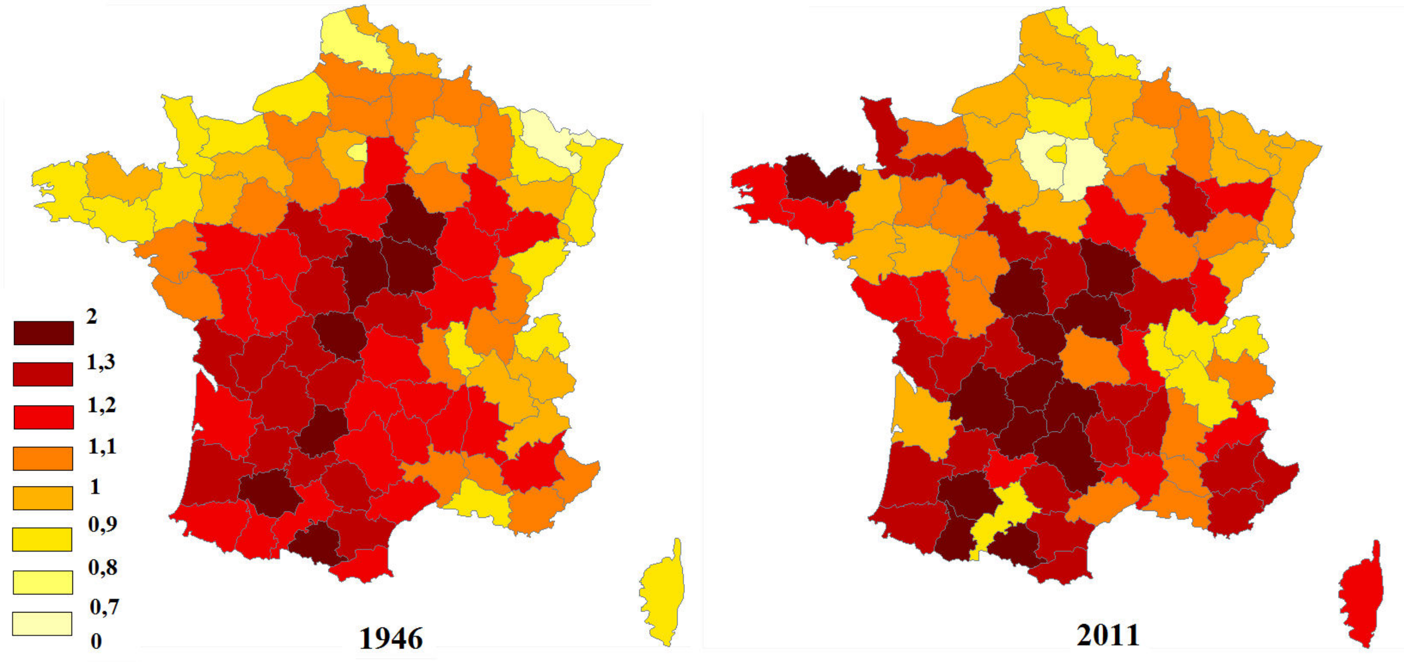
Notes: Computations based on the population of women. Ditorsion index for age x is the ratio between the population's share of age x in one *département* and the same share at the national level. These distorsion indices are smoothed over 5 years: 1946 is the mean over the 1944–1948 period. "20-30" means age 20 to 29 (30 excluded). Sample includes 90 *départements*.

distortion indices. Moreover, one can notice a remarkable situation in *Ile-de-France*: the oldest are underrepresented in the suburbs – where the values were the lowest, around 0.6 – but the situation is closer to the average in the capital and its surroundings. This phenomenon could probably be explained by the housing market: since this market is still on the rise, only the elderly who bought their flats before 1990 can stay in Paris and benefit from these amenities, while this population do not want to stay in the distant suburbs. Finally, *Bretagne* has transformed: while the elderly were underrepresented in 1946, the situation turned in 2011. For example, the share of the oldest in *Côtes-d'Armor* is 50% higher than the national one.

1.4 Conclusion

In this paper, I have linked Bonnet (2018b)'s database to census results for the period 1851–1900 in order to analyze the spatial distribution of the French population for more than 150 years. I have showed that the population was more unevenly distributed in 2014 than it was in the mid 19th century. Moreover, this increase in inequality occurred in three phases. The first, which I have named "hyper-centralization", extends from 1851 to 1900 and was due to the increase of population share living in the 10% most densely populated km^2, at the expense of all others. In particular, *Seine*'s share doubled in 50 years to reach 9.1%. The second phase, which I have named "hyper-centralization thwarted", concerns the period 1901–1968. During this period, the population was increasingly concentrated in the 10% most densely populated km^2, but this was no longer at the expense of the second, third and fourth decile. In these territories, the population grew at the

Figure 1.3.12: 65 Year-old and more distorsion indices, 1946 and 2011



Notes: Computations based on the population of women. Ditorsion index for age x is the ratio between the population's share of age x in one *département* and the same share at the national level. These distorsion indices are smoothed over 5 years: 1946 is the mean over the 1944–1948 period. Sample includes 90 *départements*.

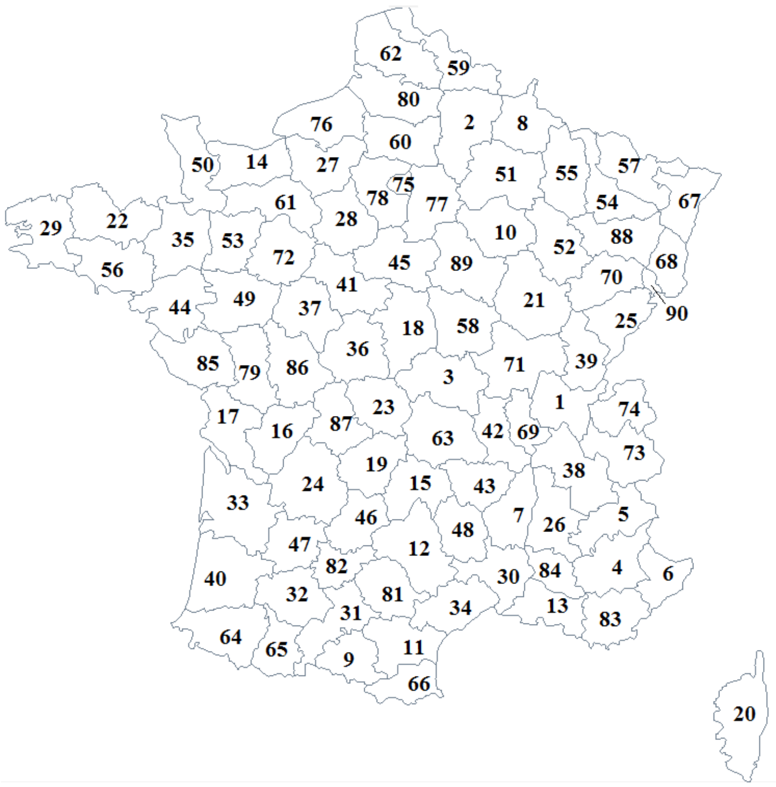
same pace than the national population. Finally, between 1968 and 2014, there was a phase of multipolarization: the share of both the most densely populated and the less densely populated areas decreased, while the share of *départements* hosting second-tier cities such as *Bordeaux, Nantes* or *Toulouse* increased. During this period, the shrinking regions of the North-East were particularly vulnerable: the share of the population living there decreased both because of migrations and birth deficit.

I have made an analysis of each age group, which highlights some specific evolutions. This is particularly the case of women aged 20 to 39, whose "hyper-centralization thwarted" has resumed in recent years: the share of the Paris region in this specific age group is increasing again. Moreover, the analysis conducted for the elderly has revealed a deconcentration phase at the end of the 19th century, due to the national homogenization of mortality conditions, in line with Bonnet et d'Albis (2018)'s results. Globally, the analysis of age structures has revealed a growing and uninterrupted territorial specialization for 50 years: at the departmental level, these age structures are increasingly different and reveal two competing France: the first composed of urban *départements* that concentrate working-age populations and the second composed of rural *départements* where the elderly are overrepresented. My cartographic analysis has deepened these results. It has showed that the geography of the oldest changed a little in 50 years: they are largely overrepresented in the South-West, in *départements* such as *Creuse, Dordogne* or *Aveyron.* Gradually, *Bretagne* joined this category because of migrations.

These demographic data make it possible to map accurately the age-specific distribution of population for more than 150 years and to highlight the main changes. This question is important because the spatial distribution of population has many public policy implications. A rapid increase in the population in some territories requires huge public investments: in the Paris region for example, the question of transport is ubiquitous. Depopulation in the central *départements* belonging to the "empty diagonal" implies a destruction of capital, both human and material: the house market is gradually depreciating, as is social cohesion. In addition, an age-specific increase may have an impact on some public policies. Thus, the increase of the youngest requires both investments in higher education and housing policies targeting singles or young couples. In contrast, the concentration of the elderly in rural areas requires a specific health policy: the issue of medical deserts is important, as is the dependency one. Thus, even if the reduction of these spatial inequalities should not be a specific objective, the reduction of imbalances through a voluntarist policy such as that conducted by the DATAR in the 1960s is important. To better understand the causes of population localization and guide public authorities in their choices, it is essential today to couple these demographic data with economic data with the same spatial framework. To go even further, the simultaneous analysis of migrations by age group, income and employment statistics over a long period would be particularly accurate.

1.5 Appendices

1.5.1 Map of the 90 French *Départements* Used to Calculate Inequalities



1 Ain
2 Aisne
3 Allier
4 Alpes (Basses)
5 Alpes (Hautes)
6 Alpes Maritimes
7 Ardèche
8 Ardennes
9 Ariège
10 Aube
11 Aude
12 Aveyron
13 Bouches du Rhone
14 Calvados
15 Cantal
16 Charente
17 Charente Maritime
18 Cher
19 Corrèze
20 Corse
21 Côte d'Or
22 Côtes du Nord
23 Creuse
24 Dordogne
25 Doubs
26 Drôme
27 Eure
28 Eure et Loir
29 Finistère
30 Gard
31 Garonne (Haute)
32 Gers
33 Gironde
34 Hérault
35 Ille et Vilaine
36 Indre
37 Indre et Loire
38 Isère
39 Jura
40 Landes
41 Loir et Cher
42 Loire
43 Loire (Haute)
44 Loire Inférieure
45 Loiret
46 Lot
47 Lot et Garonne
48 Lozère
49 Maine et Loire
50 Manche
51 Marne
52 Marne (Haute)
53 Mayenne
54 Meurthe et Moselle
55 Meuse
56 Morbihan
57 Moselle
58 Nièvre
59 Nord
60 Oise
61 Orne
62 Pas de Calais
63 Puy de Dome
64 Pyrénées (Basses)
65 Pyrénées (Hautes)
66 Pyrénées Orientales
67 Rhin (Bas)
68 Rhin (Haut)
69 Rhône
70 Saône (Haute)
71 Saône et Loire
72 Sarthe
73 Savoie
74 Savoie (Haute)
75 Seine
76 Seine Inférieure
77 Seine et Marne
78 Seine et Oise
79 Sèvres (Deux)
80 Somme
81 Tarn
82 Tarn et Garonne
83 Var
84 Vaucluse
85 Vendée
86 Vienne
87 Vienne (Haute)
88 Vosges
89 Yonne
90 Territoire de Belfort

Notes: Numbers used in Bonnet (2018b)'s database. Corse is unified in this classification.

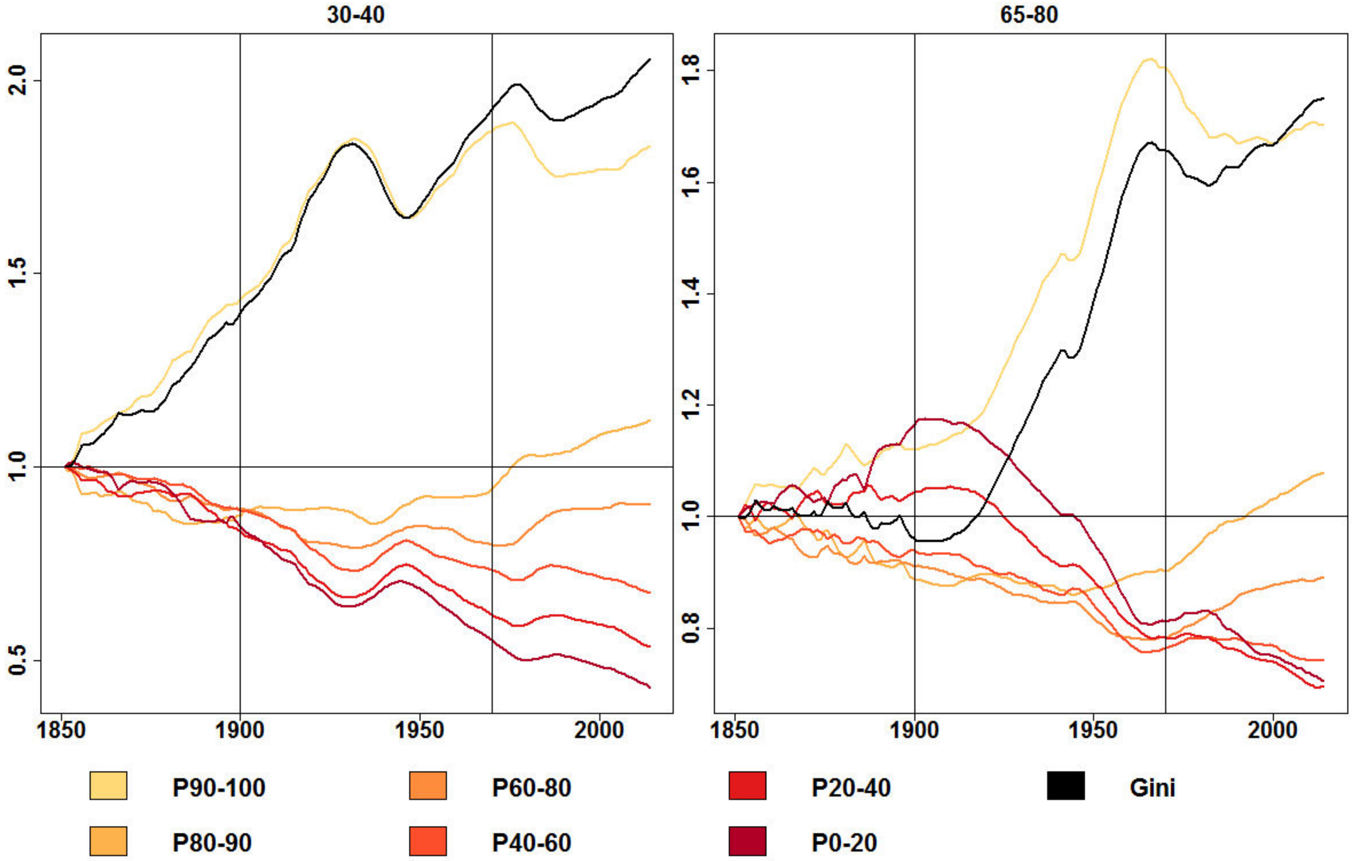
1.5.2 Spatial Distribution of Men Aged 30 to 39 and 65 to 79: Inequality indices

Figure 1.5.1 shows how the spatial distribution of men aged 30 to 39 and those aged 65 to 79 evolved for 150 years. One can see that the changes are similar to those of the younger women, with the difference that the Second World War led to a larger decline in the inequality indicator. As for the older ones, one can note that the share of the first decile in the total population is starting to increase since the end of the 2000s, which was not the case for women.

1.5.3 Evolution of the Distribution of Population by Age Group : Decomposition Method

I seek to decompose the evolution of the distribution of population by age group between the background process (namely the evolution of total population densities) and the specific process (namely the share of the age group in the departemental population). Formally, let call d_t^a the population density of age a at date t in one *département*, and S the area. One have:

Figure 1.5.1: Inequalities of Spatial population density of men by age group, 1851-2014



Notes: Computations based on the population of men. P90-100 refers to the share of national population who lived in the 10% of km^2 with the highest density values. All inequality indicators are weighted by km^2 and normalized by 1851 values. "30-40" means age 30 to 39 (40 excluded). Sample includes 90 *départements*.

$$d_t^a = \frac{\theta_t^a P_t}{S}, \tag{1.5.1}$$

with θ_t^a the share of population aged a in the total of population P_t. The area is fixed regardless of age and year for a *département*. The sequence of the inequality index Θ^a relative to age a for the T years studied can be written as follows:

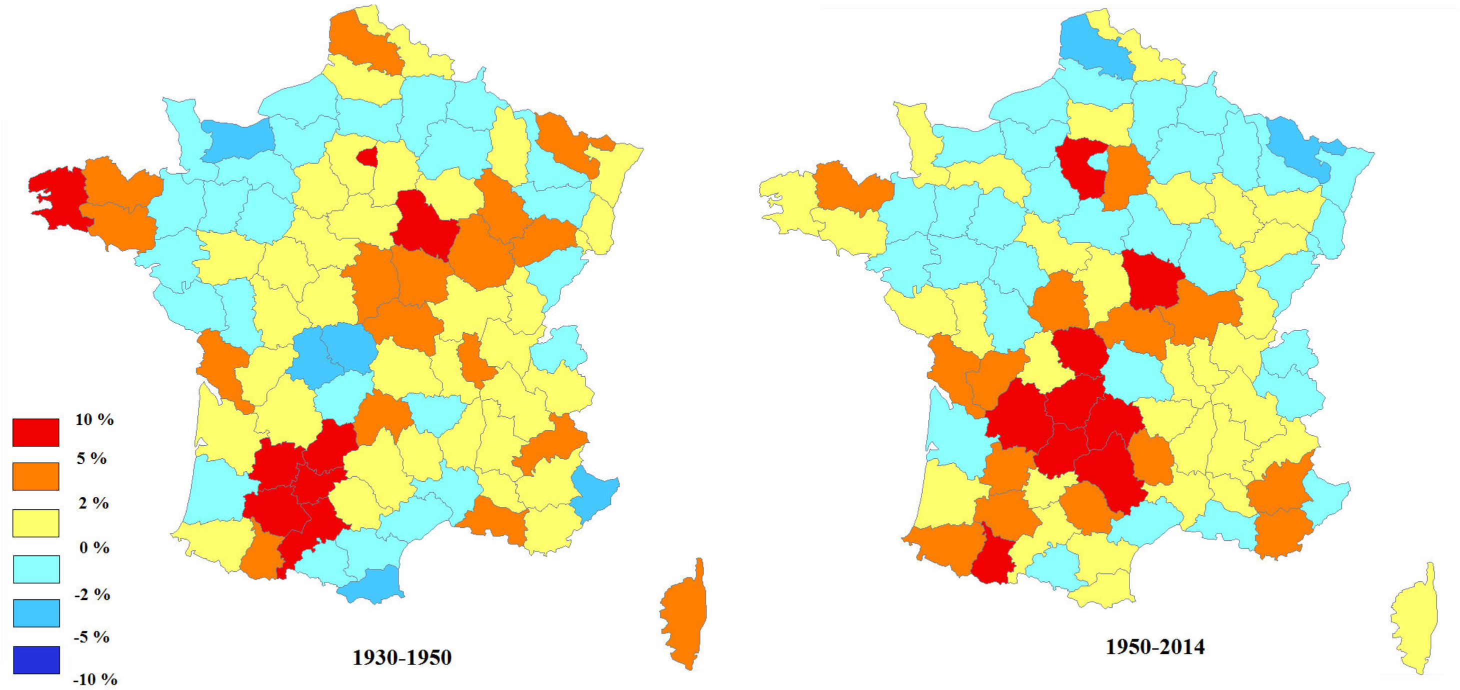
$$\Theta^a = \Theta^a \left(\boldsymbol{\theta^a}, \mathbf{P} \right), \tag{1.5.2}$$

where $\boldsymbol{\theta^a}$ is the (N, T) matrix of departmental age structures for each year and $\mathbf{P}$ is the (N, T) matrix of the departmental total population for each year. The variation of Θ^a comes either from the evolution of θ_t^a, specific to the age group a, or from the evolution of P_t, common to all age groups. To assess whether departmental age structures contribute to the increase of inequalities, I fix departmental populations at their initial level and allow θ_t^a to vary.

1.5.4 Departmental Contributions in KLD_{Nat} Evolution

Thanks to its additivity property, the KLD_{Nat} can be decomposed to know the weight of each *département* in its evolution between two dates. Figure 1.5.2 shows the result of this decomposition for the periods 1930–1950 (during which the indicator fell by 30%) and 1950–2014 (during which the indicator increased by 80%). For the first, one can identify four geographical areas explaining the decline: *Seine*, *Bretagne*, *Yonne* and its neighboring *départements*, the South-West along the Garonne Valley. For the second, I identify four geographical zones explaining the rise: *Seine-et-Oise*, *Hautes-Pyrénées*, *Nièvre* and a large region around *Corrèze* in the South-West.

Figure 1.5.2: DEPARTEMENTAL CONTRIBUTIONS IN KLD_{Nat} VARIATIONS, 1930–1950 AND 1950–2014



Notes: Computations based on the population of women. Since $KLD_{Nat}$is the addition of departmental dissimilarities, one can calculate the contribution of each *département* in the total of variation. The variation between 1930 and 1950 is negative: a positive contribution means that the *département* is closer to the national age structure. On the opposite, the variation between 1930 and 1950 is positive. Sample includes 90 *départements*.

1.5.5 Distorsion indices in 1856, 1896, 1946 and 2011

Distorsion indices are the ratio between the share of the age group in the departmental population and the share of this age group at the national level. The distortion indices are smoothed by 5-year period. The values are calculated for 4 years among the 150 years of the study in order to extend the map analysis.

	Women aged 20 to 29				Women aged 65 and over			
	1856	1896	1946	2011	1856	1896	1946	2011
Ain	98%	94%	96%	89%	84%	105%	108%	88%
Aisne	95%	91%	100%	92%	129%	125%	103%	100%
Allier	103%	100%	89%	74%	74%	95%	128%	137%
Alpes (Basses)	98%	90%	95%	75%	85%	103%	117%	125%
Alpes (Hautes)	98%	89%	102%	75%	64%	75%	95%	116%
Alpes Maritimes	106%	120%	94%	87%	94%	81%	108%	127%
Ardèche	102%	93%	92%	74%	103%	96%	114%	120%
Ardennes	94%	93%	96%	88%	127%	125%	105%	102%
Ariège	101%	89%	85%	71%	86%	129%	147%	131%
Aube	96%	88%	95%	91%	144%	145%	108%	108%
Aude	93%	98%	92%	76%	99%	108%	125%	125%
Aveyron	96%	90%	94%	70%	92%	103%	116%	141%
Bouches du Rhone	94%	114%	107%	102%	116%	79%	90%	101%
Calvados	93%	93%	104%	99%	160%	130%	87%	103%
Cantal	103%	94%	94%	70%	95%	113%	117%	143%
Charente	90%	90%	92%	77%	110%	135%	124%	125%
Charente Maritime	97%	93%	97%	76%	124%	126%	125%	130%
Cher	97%	92%	89%	78%	67%	94%	134%	125%
Corrèze	98%	95%	96%	73%	85%	87%	123%	141%
Corse	117%	111%	116%	92%	58%	66%	89%	116%
Côte d'Or	91%	89%	96%	110%	142%	140%	116%	102%
Côtes du Nord	99%	93%	102%	73%	88%	86%	94%	131%
Creuse	102%	96%	82%	65%	75%	114%	154%	154%
Dordogne	98%	96%	95%	69%	90%	100%	123%	139%
Doubs	100%	97%	104%	108%	90%	86%	87%	96%
Drôme	115%	101%	97%	86%	95%	103%	111%	107%
Eure	87%	88%	97%	90%	162%	144%	100%	91%
Eure et Loir	84%	89%	96%	89%	138%	129%	105%	97%
Finistère	105%	100%	106%	84%	63%	65%	85%	118%
Gard	101%	105%	96%	85%	96%	99%	108%	110%
Garonne (Haute)	101%	99%	102%	124%	101%	119%	114%	86%
Gers	90%	90%	92%	65%	125%	130%	130%	135%
Gironde	118%	111%	99%	107%	89%	111%	116%	98%
Hérault	105%	108%	98%	107%	99%	104%	117%	105%
Ille et Vilaine	97%	103%	105%	112%	63%	75%	86%	91%
Indre	105%	94%	90%	71%	68%	102%	125%	137%
Indre et Loire	91%	93%	95%	102%	118%	128%	110%	106%
Isère	105%	94%	99%	102%	64%	109%	98%	90%
Jura	94%	95%	97%	80%	102%	102%	108%	114%
Landes	96%	107%	90%	72%	96%	65%	128%	120%
Loir et Cher	93%	89%	90%	77%	99%	113%	121%	122%
Loire	104%	108%	96%	93%	66%	76%	101%	114%

	Women aged 20 to 30				Women aged 65 and over			
	1856	1896	1946	2011	1856	1896	1946	2011
Loire (Haute)	101%	95%	89%	72%	74%	93%	119%	120%
Loire Inférieure	98%	103%	100%	102%	94%	90%	101%	94%
Loiret	95%	90%	95%	97%	100%	114%	116%	99%
Lot	95%	110%	88%	62%	117%	136%	141%	142%
Lot et Garonne	96%	96%	95%	79%	131%	127%	122%	129%
Lozère	96%	86%	93%	77%	86%	85%	113%	127%
Maine et Loire	99%	98%	97%	101%	112%	110%	111%	98%
Manche	90%	96%	103%	79%	148%	101%	88%	123%
Marne	94%	95%	102%	109%	121%	118%	95%	94%
Marne (Haute)	93%	83%	95%	84%	135%	150%	113%	121%
Mayenne	97%	89%	99%	83%	104%	102%	94%	110%
Meurthe et Moselle	105%	99%	107%	113%	109%	102%	81%	98%
Meuse	89%	84%	97%	86%	130%	137%	108%	109%
Morbihan	112%	102%	101%	78%	67%	79%	87%	119%
Moselle	88%		111%	101%	93%		68%	97%
Nièvre	96%	87%	86%	72%	75%	107%	139%	143%
Nord	95%	111%	101%	115%	98%	77%	93%	86%
Oise	88%	95%	97%	99%	138%	121%	103%	80%
Orne	90%	90%	98%	80%	137%	141%	99%	124%
Pas de Calais	96%	98%	110%	99%	111%	77%	75%	94%
Puy de Dome	99%	98%	96%	100%	118%	113%	118%	109%
Pyrénées (Basses)	97%	100%	96%	84%	102%	106%	111%	123%
Pyrénées (Hautes)	97%	98%	97%	72%	98%	130%	119%	137%
Pyrénées Orientales	98%	104%	99%	82%	73%	80%	113%	127%
Rhin (Bas)	95%		103%	111%	73%		81%	90%
Rhin (Haut)	123%		97%	95%	66%		89%	95%
Rhône	121%	117%	102%	126%	73%	78%	90%	89%
Saône (Haute)	98%	85%	91%	80%	115%	126%	120%	107%
Saône et Loire	103%	106%	93%	76%	89%	82%	119%	126%
Sarthe	91%	89%	98%	88%	108%	132%	100%	108%
Savoie	99%	93%	100%	92%	63%	97%	98%	100%
Savoie (Haute)	101%	96%	107%	97%	66%	77%	87%	84%
Seine	132%	128%	110%	124%	75%	68%	78%	81%
Seine Inférieure	107%	99%	107%	105%	121%	96%	83%	98%
Seine et Marne	86%	91%	90%	109%	134%	125%	113%	69%
Seine et Oise	88%	100%	100%	119%	139%	102%	91%	65%
Sèvres (Deux)	95%	96%	95%	80%	97%	101%	110%	116%
Somme	92%	90%	97%	104%	123%	128%	110%	97%
Tarn	96%	97%	92%	75%	94%	118%	124%	128%
Tarn et Garonne	101%	94%	93%	78%	132%	138%	125%	113%
Var	104%	110%	100%	80%	110%	110%	105%	127%
Vaucluse	99%	98%	97%	89%	98%	117%	106%	107%
Vendée	100%	97%	95%	81%	74%	83%	106%	115%
Vienne	98%	95%	95%	107%	97%	107%	116%	107%
Vienne (Haute)	106%	100%	95%	94%	61%	80%	121%	122%
Vosges	101%	97%	96%	82%	92%	99%	100%	114%
Yonne	89%	81%	85%	80%	127%	154%	140%	118%
Belfort	125%	105%	98%	100%	80%	75%	92%	97%

Chapter 2

Spatial Inequalities in French Life Expectancy, 1806–2014

Abstract

This article analyzes spatial inequalities in mortality in France over the past 200 years. Using a new database, we identify a period from 1881 to 1980 when inequalities rapidly shrank while life expectancy rose. This century of convergence between the parts of France was mainly due to the fall in infant mortality. Since 1980, spatial inequalities have levelled out or occasionally widened, due mainly to differences in life expectancy among the elderly. The geography of mortality also changed radically during the century of convergence. Whereas in the 19th century high mortality occurred mainly in larger cities and along a line from north-west to south-east France, it is now concentrated in the north, and *Paris* and *Lyon* currently enjoy an urban advantage.[1]

[1]This paper is co-written with Hippolyte d'Albis.

2.1 Introduction

Since the French Revolution, France has been subdivided into *départements*, geographical units whose boundaries have changed little over time. In 2014, the difference in life expectancy at birth between women living in the top *département* (*Aisne*) and the bottom one (*Savoie*) was 3 years 10 months. This gap may seem large, especially for a country where equality figures in the national motto, and public health and welfare policies are among the most extensive in the world. This article puts that gap into perspective by following the development of life expectancy inequalities between French *départements* over the last two centuries.

Most research on spatial inequalities of mortality concerns recent periods and reveals either a levelling out or an increase in inequality. In Germany, for example, Kibele's (2012) regional mortality analysis for 1990-2006 shows that the inequality index has barely varied since 1995. This is due to two countervailing trends: increased inequality in the West German Länder, and convergence to the national mean in the Länder of the former GDR.[2] In the United Kingdom, Illsley and Le Grand (1993) include a further dimension with an analysis of spatial inequalities at various ages, showing that inequalities of mortality have reduced among younger people and increased among the oldest. In France, Daguet (2006), Barbieri (2013) and Breton et al. (2017) also reveal a stabilisation in the widest gap in female life expectancy between *départements* since 2000 and a widening of the male gap since 1995.

These recent developments differ greatly from those in the historical record. In France, Bonneuil (1997) reconstitutes *département* life expectancies for 1806-1906, showing that the standard deviation of female life expectancy at birth fell throughout the 19th century, with the notable exception of 1851 and 1870. Vallin and Meslé (2005) use his reconstitution to show that *département* convergence was rapid in the closing decades of the 19th century. This pioneering work, however, only marginally examined the development of spatial inequalities. Indeed, the long-term historical trends could not be analyzed because of the lack of data for 1906-1975.

This article uses a new database, presented in Bonnet (2018b), which reconstitutes annual life tables for each French *département* and both sexes for 1901-2014. This period includes the two World Wars, for which the author uses data from new sources on civilian and military deaths and those in deportation. Combining this new base with Bonneuil's (1997), we have analyzed the development of life expectancy inequalities between French *départements* since 1806.

Our first contribution is to characterize the development of inequalities over time and date the three main phases identified. The period from 1806 to 1880 alternates divergence and convergence between regions and rises and falls in national life expectancy. The 1881-1980 period we call the "century of convergence", a fortunate time when inequality shrank and life expectancy rose. Both occurred regularly except for the years of the two World Wars. The third period, 1981-2014, returns to an alternation of convergence and divergence along with a general rise in life expectancy. This alternation also figures in Vallin and Meslé's research (2005), where they hypothesize and identify

[2] See also Brown and Rees (2006) for Yorkshire, Ezzati et al. (2008) for US counties, Joseph et al. (2009), for Canada, Gächter and Theurl (2011) for Austria, Barbieri and Ouellette (2012) for Canada and the United States, and Janssen et al. (2016) for the Netherlands.

"Matthew effect" years of divergence when rising life expectancy occurs mainly in those areas where longevity is already highest. Age-group analysis also shows that developments in infant mortality account for most of the reduction in spatial inequalities over the century of convergence. At present, it is basically inequalities among the oldest people that account for spatial inequalities.

Our second contribution is a geographical analysis of areas of high mortality. We take a number of areas similar in life expectancy profile that have experienced notable change over these two centuries. One striking example is the northern *départements* in France: they were well ahead in the 19th century but are now well below the national mean. Conversely, the Seine *département*, containing Paris, long suffered an "urban penalty" (Haines, 2001) which has become an "urban advantage" since the 1930s.

Section 2 presents the data and indicators used and describes the main concepts used. Section 3 presents our results and Section 4 concludes.

2.2 Data and Methods

We analyze the developments in spatial inequalities of life expectancy in France since 1806. The indicator most commonly used in the literature is life expectancy at birth, calculated from current life tables. Other indicators may also be used. Edwards and Tuljapurkar (2005) use life expectancy at age 10 so as to eliminate the effects of infant mortality. Standardised (Brown and Rees, 2006) or raw mortality rates are also sometimes used in order to eliminate differences between population age structures. In this article, we use life expectancy at various ages that we have calculated from current life tables for French *départements*. We do not analyze inequalities within *départements* but focus on inequalities between *départements*. So life expectancy at a given age shows the number of years of life left for an average individual living in a given *département*.

2.2.1 Data Sources

We use data from Bonneuil (1997) for 1806-1900 and from Bonnet (2018b) for 1901-2014.

Bonneuil (1997) calculates current life tables for women for each French *département*. These tables are given in five-year age-groups from 0 to 85, and for five-year periods in 1806-1906. Note that deaths and censuses by age-groups are only available from 1856. For 1806-1855, Bonneuil estimates life expectancies using the Ledermann model. He takes life tables and age pyramids for 1856-1906 and total births and deaths for 1806-1906. This gives differing life expectancies at birth for the *départements* but life tables with an identical internal structure: differences in death rates by age are frozen at the 1856 value. Given this limitation, we only use life expectancies at birth for 1806-1856 and entire life tables for 1856-1901.

For 1901-2014 we have life expectancies, mortality rates and population figures at 1 January of each year, for women and men, calculated by Bonnet (2018b), mainly using the Human Mortality Database protocol. The raw data he compiled came from the archives of the French statistical agencies. This database contains previously unavailable information on mortality during the first half of the 20th century. It complements the *département* database in Daguet (2006), which covers the census years during the 1954-1999 period.

In order to aggregate these databases, an analysis was made of the differences in life expectancy at birth between Bonnet (2018b) and Bonneuil (1997) and Daguet (2006). This involved Bonneuil's data on 1901-1905, where *département* life expectancies are available in both sources. The comparison with Daguet (2006) concerned 1954-1999. Table 2.1 shows the differences observed for the two periods. For 1901-1905, these differences are quite large, and life expectancies are on average lower in Bonneuil. For the later period, the differences are small. Life expectancies are slightly lower in Daguet.

Table 2.1: DIFFERENCES IN LIFE EXPECTANCIES AT BIRTH (IN %): BONNEUIL-BONNET 1901–1906; DAGUET-BONNET 1954–1999

	Men			Women		
	Quart. 1	Med.	Quart. 3	Quart. 1	Med.	Quart. 3
1901–1905				0.49	3.34	6.05
1954	0.18	0.65	1	0.54	0.84	1.34
1962	0	0.4	0.72	-0.01	0.37	0.68
1968	0.17	0.38	0.73	-0.02	0.33	0.78
1975	-0.17	0.15	0.5	-0.11	0.19	0.47
1982	0.01	0.27	0.59	0.04	0.21	0.5
1990	0.09	0.31	0.55	0.21	0.4	0.62
1999	0.22	0.49	0.73	0.47	0.66	0.99

Notes: 25% of the differences are below the threshold "Quart. 1". "Med." is the usual median. Sample includes 90 *départements*.

The two sources used for this study do not present the same variables for mortality rates by age, because Bonneuil (1997) does not give mortality rates above age 90. The Bonnet (2018b) database shows than in 1901-1905 the number of survivors beyond 90 was small (from 0.2% to 2%, depending on year and *département*), and the mortality rates for ages 85 to 90 vary little and are close to 75%. Consequently a mortality rate of 100% was applied to the 90-95 age-group for the years before 1901.

2.2.2 Geographical Scope

This article examines inequalities in life expectancy between French metropolitan *départements*. The overseas *départements* have not been included because their demographic statistics are much more recent and cannot be extrapolated backwards. In this article, we use the term "national" to describe the situation in metropolitan France.

The number and boundaries of the French metropolitan *départements* have generally varied little since they were created in 1790. What few modifications there have been are due to changes in France's eastern border and a recent reorganization of the Paris region. In order to have a consistent comparison over time, we have applied the classification of 90 *départements* valid in 1967 (see map in Appendix A1). This required reconstituting the life tables for some *départements* that were absent in certain census years. We used an adjacent *département*, called a reference *département*.

The life tables of the reference *départements* were used to re-estimate those of the missing *départements*. The fit between the tables of each pair of *départements* was analyzed for the closest year for which the data are available. Then specific ratios were calculated and applied to the life tables of the reference *département* for the missing years. In formal terms, $l(x)$ is the number of survivors at age x, indexed by r for the reference *département* or m for the missing *département* and $t*$ for the reference year or t for the estimated year.

The following formula is applied:

$$\frac{100,000 - l_m^t(x)}{100,000 - l_m^r(x)} = \frac{100,000 - l_m^{t*}(x)}{100,000 - l_m^{r*}(x)} \tag{2.2.1}$$

from which is calculated the number of survivors in each *département* that is lacking a year's figures. This is used to calculate the corresponding life expectancy.

Table 2.2 shows the 10 *départements* for which data are lacking and the reference *département* used for each one.

Table 2.2: DEPARTEMENTS WITH MISSING DATA

Département	Period with missing data	Reference *départements*
Alpes-Maritimes	1806–1856	*Var*
Meurthe-et-Moselle	1806–1866	*Meuse*
Savoie	1806–1856	*Ain*
Haute-Savoie	1806–1856	*Ain*
Territoire de Belfort	1806–1866	*Meuse*
Moselle	1866–1920	*Meuse*
Bas-Rhin	1866–1920	*Meuse*
Haut-Rhin	1866–1920	*Meuse*
Seine	1969–2014	(*)
Seine-et-Oise	1969–2014	(*)

Notes: Reference *départements* are *départements* used to estimates values for missing *départements.*
(*) Sum of *Essonne, Hauts-de-Seine, Seine-Saint-Denis, Val-de-Marne, Val d'Oise, Paris, Yvelines.*

Three *départements* in south-eastern France (*Alpes-Maritimes, Haute-Savoie* and *Savoie*) were only created in 1860 when they were transferred to France. So for *Alpes-Maritimes*, with *Var* as its reference *département*, ratios for mortality rates at each age in the two *départements* for 1861 and 1866 were calculated and then applied to the *Var* 1856 life tables to estimate those of *Alpes-Maritimes* for that year.

In the case of the two *départements* created in 1871 from the territory not annexed by Germany (*Meurthe-et-Moselle* and *Territoire de Belfort*), the same procedure was used to reconstitute data up to 1866. For the three *départements* that were German until the First World War (*Bas-Rhin, Haut-Rhin* and *Moselle*), the same was done up to 1900. After that date we have the annual data in Bonnet (2018b) for deaths and population by age. The same assumption is made of stable ratios between missing *département* and reference *département*, but only for those variables. Then the corresponding life expectancies were calculated.

As a result of their large populations, the two *départements* of the Paris region (*Seine* and *Seine-et-Oise*) were divided into seven in 1968. The deaths and population by age in the new *départements* were allocated to the old *départements* in the proportions observed in 1968, the year for which both sets of data are available. This ensures consistency in the data since the totals do not change. The corresponding life expectancies were then calculated.

None of these reconstituted data have any effect on the results given below. The results remain unchanged when the *départements* that required data reconstitution are excluded.

2.2.3 Indicators of Inequality

There are a large number of indicators that measure inequality in general and inequality in mortality in particular. Mackenbach and Kunst (1987) show that each indicator provides different information. Indicators based on extreme points (gap or ratio between the best-placed and worst-placed individuals) or an interquantile range (gap or ratio between the best-placed $x\%$ and $(100-x)\%$ worst-placed) differ from indicators that use the entire distribution such as the Gini coefficient or the concentration index. Dissimilarity indices show how much would need to be redistributed among groups for mortality to be the same for all. For example, d'Albis et al. (2014) use the Kullback-Leibler divergence to analyze differences in mortality rates at each age.

In this article two main types of indicator are used. One simple indicator is the Gini coefficient. Another series of indicators are used to analyze specific sections of the *département* life expectancy distribution. By aggregating all the *département* life expectancies, the sum of "total years lived" is defined and a proportion is calculated for a number of *départements* ranked by life expectancy. So the top decile (designated P90-100) contains the 10% of *départements* with the highest life expectancy, the second decile (P80-90) the next 10% and so on. To obtain indicators of inequality similar to those used to analyze spatial disparities, these proportions are compared with what they would be in an even distribution. If the proportion of total years lived by the top 10% is 15% (as opposed to 10% in an even distribution), then the inequality indicator is 50%. This means that the 10% of *départements* with the highest life expectancy have 50% more years of life than they would have if these years were evenly distributed across France. This calculation makes it possible to have homogeneous values for all the inequality indicators, each for a specific interval of the distribution curve.

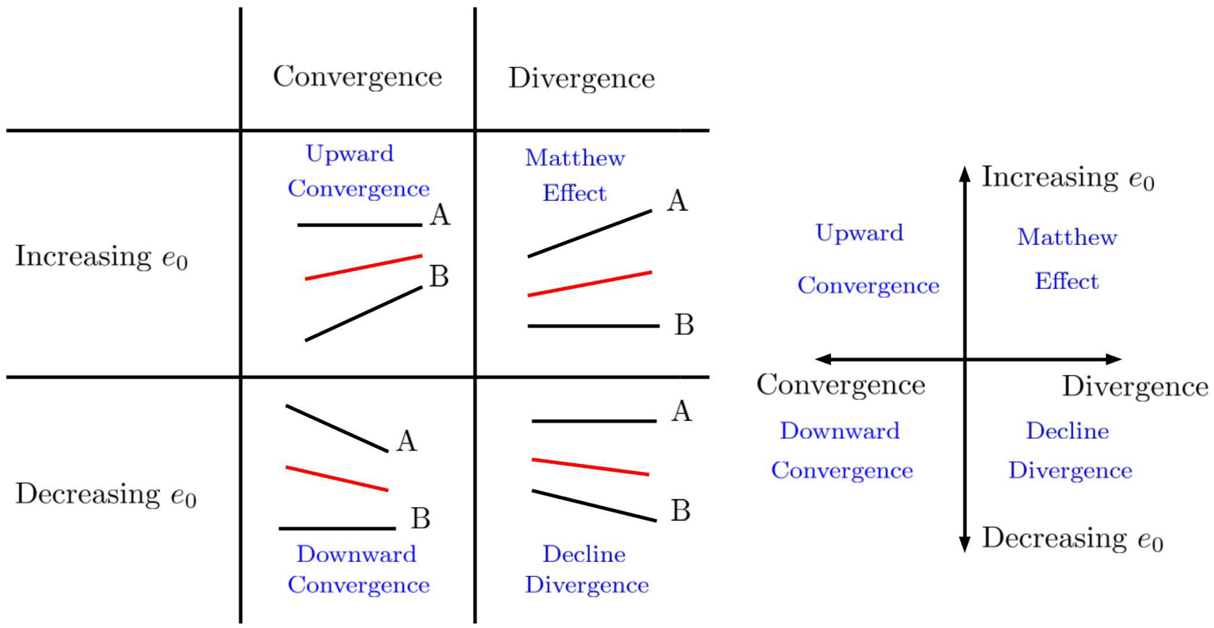
The question of whether to weight the *départements* in calculating inequalities is not an easy one. French *départements* have relatively similar surface areas but widely varying population densities. In 1901, for example, the estimated density in *Basses-Alpes* (now *Alpes-de-Haute-Provence*) was 16 people per square kilometer, compared with 7,500 in Seine. Weighting *départements* by their population assumes that what matters is the general welfare of individuals, with no emphasis on less populated areas. Conversely, not to apply weighting can be justified for the purposes of public policy, where territorial and political issues prevail. In this article, we have chosen to give unweighted indicators, with each decile comprising 9 *départements*, and the *département* mean does not correspond to the national mean. Weighted indicators are only given if notable differences emerge.

2.2.4 Convergence Indicators

Our analysis of convergence is essentially graphical, showing the variation of these inequality indicators over time. Any reduction in inequality corresponds to a convergence between *départements* and an increase to a divergence. These graphical analyzes could be supplemented by econometric sigma-convergence analyses (Gächter and Theurl, 2011, Janssen et al., 2016), but these estimates are usually unnecessary. Our approach differs, however, from beta-convergence analyses, where a regression is run between the variation in mortality between two dates and the initial mortality figures. This sort of regression depends too much on the choice of start date.

Any analysis of convergence that does not take account of demographic changes during the period would be simplistic and lead to absurdities. For example, while it is desirable that the *départements* with the lowest life expectancies should converge by catching up, this is not so if the top *départements* reduce their life expectancy. An illustration would be two *départements*, A and B, where A has the highest and B the lowest life expectancy. There are four possible scenarios for changes in A and B's life expectancies. Figure 2.2.1 presents them by changes in inequality (convergence and divergence) and changes in average life expectancy (rising or falling).

Figure 2.2.1: Divergence and convergence periods: theoretical classification



Notes: Convergence is a decrease in lifespan inequalities between *départements.* Increasing in e_0 is an increase of national lifespan. National lifespan is the lifespan mean of our sample. Sample includes 90 *départements*.

The first convergence scenario ("upward convergence") shows a catch-up phase in which B reduces the gap with A by improving its life expectancy. This sort of convergence sees greater consistency across France and a general rise in life expectancy. The second scenario ("downward convergence") involves a fall in average life expectancy. This would occur if A had a falling life expectancy. This type of convergence is observed particularly in war time.

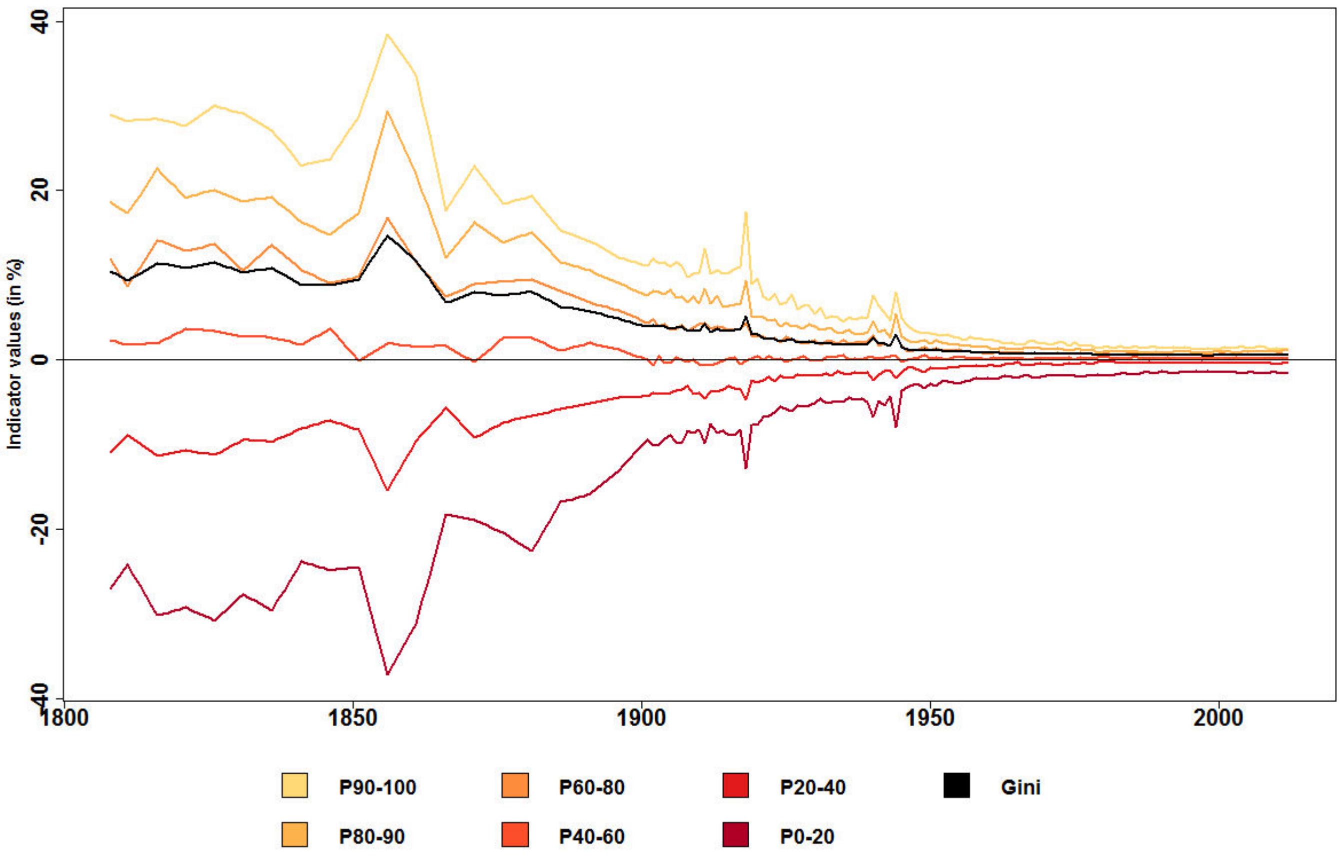
The first divergence scenario sees a greater rise in A's life expectancy than in B's. This is "Matthew effect" divergence, by which "to him who has will more be given" (Merton, 1968). It may be that advances in technology or medicine are first enjoyed by the more favoured, who thus widen their advance, but the less favoured may catch up later. Vallin and Meslé (2005) describe this alternation between phases of divergence and convergence. The second divergence scenario ("decline divergence") sees increased inequality and falling average life expectancy. This may occur in particular during epidemics.

2.3 Results

2.3.1 The Three Phases in Reduction of Spatial Inequalities of Life Expectancy

First the variations in spatial inequalities of life expectancy since 1806 are analyzed. We use the Gini index and the relative share of each decile in the distribution. Figure 2.3.1 shows these indicators over time, calculated for women with no population-weighting of *départements*. This is the only graph that can be calculated for the whole period, because from 1806 to 1900 no data are available for *département* male life expectancy or female population.

Figure 2.3.1: SPATIAL MORTALITY INEQUALITIES FOR WOMEN, 1806–2014



Notes: P90-100 means the share of lifespan lived by the 10% of *départements* with the highest values (compared with a uniform lifespan for all *départements*). All inequality indicators are non-weighted by population, for women. Sample includes 90 *départements*.

Figure 2.3.1 shows that spatial inequalities of life expectancy have fallen sharply since 1806. The

Gini index has fallen from 0.105 to 0.005 in 2014. There is no clear trend in the first half of the 19th century and the fall starts in around 1860 with brief interruptions during the three wars between France and Germany. The absolute range in life expectancy between the top and bottom *départements* shrank from 39 years 1 month in 1856 to 3 years 10 months in 2014. The indicators calculated for the various distribution deciles show that the catching up process was highly effective for the 18 *départements* with the lowest life expectancy (P0-20 in Figure 2.3.1). In 1856, they had nearly 40% fewer life years than in an ideal even distribution, compared with 17% in the early 20th century and 1.5% in 2014.

From 1901, the data from Bonnet (2018b) can be used to test the robustness of these indicators by analysing spatial inequalities of life expectancy for men and weighting the *départements* by population. The results for the Gini index are given in Appendix A2. The values of the indicators differ but the general trend is unchanged.

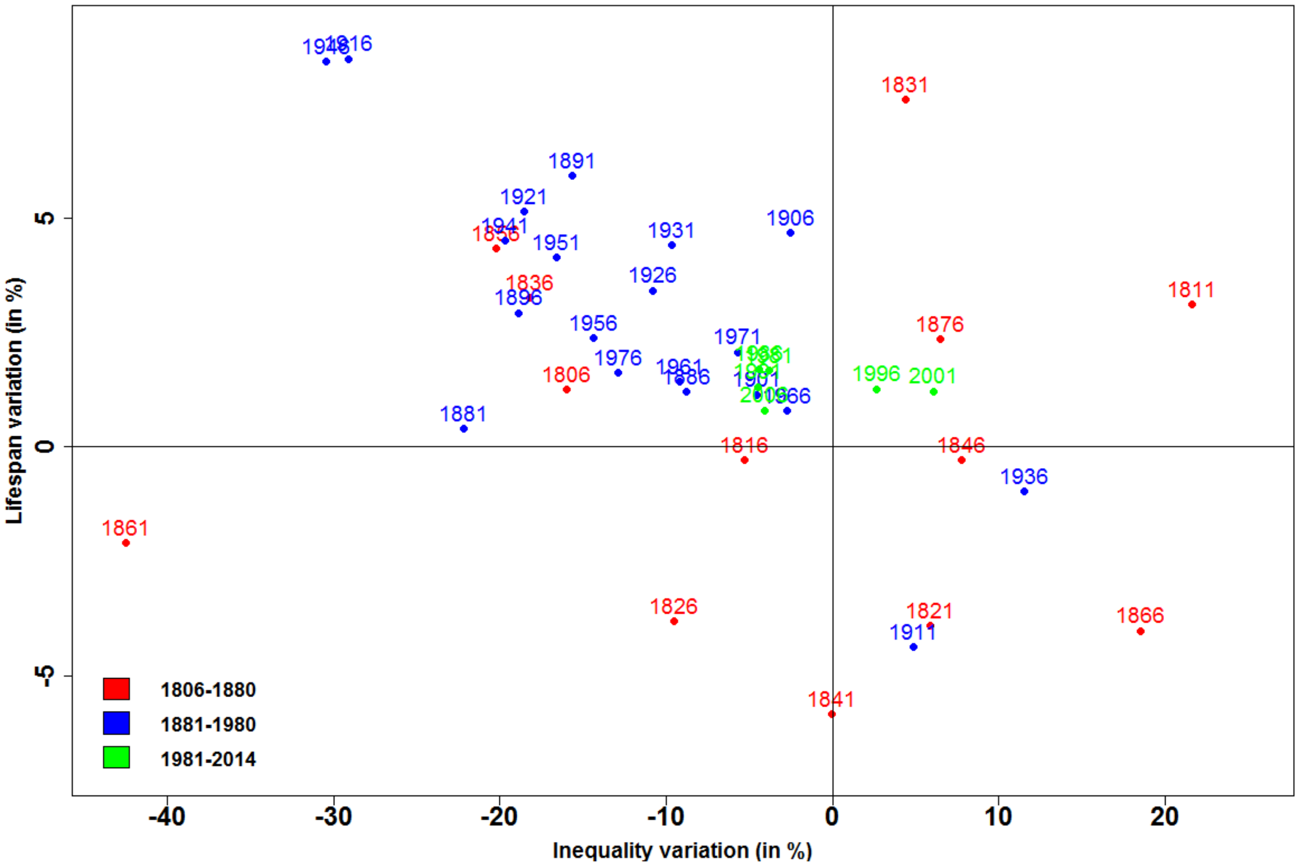
As explained above, it would be simplistic to analyze variations in inequalities of life expectancy without taking account of the development mean life expectancy in France as a whole. For that reason, variations in the Gini index and life expectancy in France were calculated for each five-year period from 1806 to 2014.[3] The pairs of points obtained for each period are plotted on a four-quadrant diagram (Figure 2.3.2). Quadrant II (upper left), for example, contains periods where there is both a reduction in inequalities and a rise in life expectancy.

Figure 2.3.2 can be used to identify three phases in the reduction of spatial inequalities of life expectancy. The first phase, 1806-1880, displays non-convergence. In Figure 2.3.2 the points from this phase are scattered across the four quadrants. During this period mean life expectancy and inequality went up and down with no real trend perceptible. This came to an end in about 1880, after which life expectancy increased continuously, except for the two World Wars. This marks the second phase, 1881-1980, which may be called the century of convergence. In Figure 2.3.2 all these points lie in Quadrant II, except for those affected by the two World Wars, 1911 and 1936. The temporary rise in inequalities during the wars is due to the highly varied exposure of regions to the fighting. For example, in 1944, 40% of deaths in Calvados were due to the bombing associated with the Normandy landings (Bonnet, 2018a). This century of convergence generally involved a virtuous process whereby inequalities shrink and at the same time mean life expectancy increases. Note that this reduction in inequality began nearly 50 years before the introduction of public insurance systems for illness, old age and death . This noteworthy convergence came to an end in about 1980. There then followed a phase during which mean life expectancy continued to rise while inequalities stopped shrinking. Indeed they even widened from 1996 to 2005, a case of “Matthew effect” divergence (Quadrant I of Figure 2.3.2).

Finer analysis of variations in life expectancy inequality since 1980 shows a specific pattern in the least-favoured *départements*. Figure 2.3.3 repeats Figure 2.3.1 with a focus on 1980-2014. It can be seen that inequalities are indeed much smaller than before. The mean life expectancy in the bottom two deciles is 1.8% less than an even distribution, and the top decile mean is 1.3% higher. However, the curve showing the relative position of the bottom two deciles, which rises until 1995, levels out

[3]Point 1806, for example, represents the variations between the two five-year periods 1806-1810 and 1811-1815.

Figure 2.3.2: The three phases of the convergence process, 1806–2014



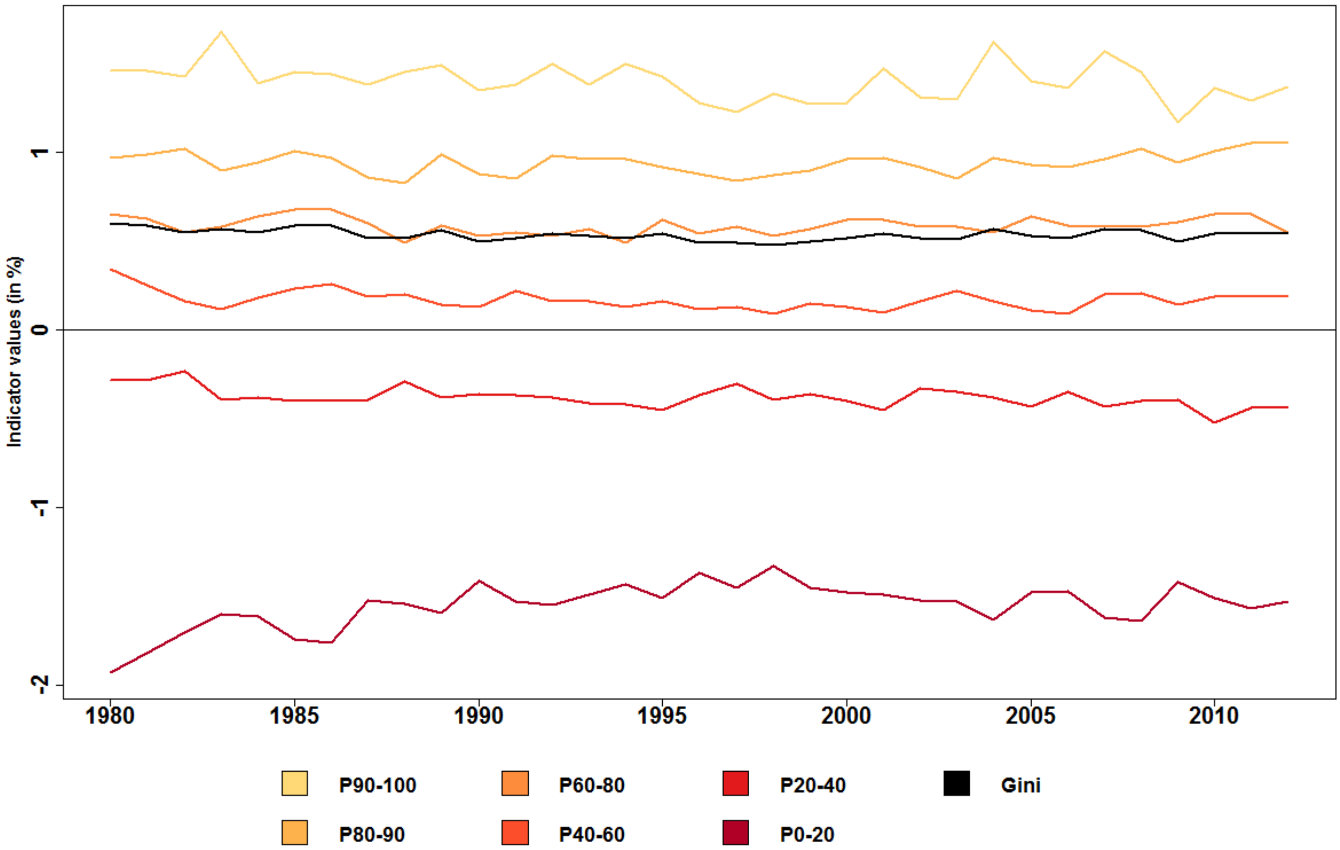
Notes: Variation of the national lifespan for year 1806 is the variation of lifespan between the 1806–1810 quinquennial period and the 1811–1815 one. We used the quadriennal period 2011–2014 for the 2006 point. Gini indicator is non-weighted by population, for women. Sample includes 90 *départements*.

and even falls from 1996 to 2005.

The evolution of French spatial inequalities that we reveal can be compared with the evolution of inequalities at the international level and between American states. In the first case, we used data from the United Nations Population Division, available for the period 1950-2015, and analyzed the inequalities between the 201 geographical units for the female population. The results are the same depending on whether the geographical units are weighted or not by the population: inequalities decline over the entire period. These results are different from the French case: there has been no rise in inequality over the recent period. In the second case, we used data from the US Mortality Database[4], available for the period 1959-2015, and analyzed the inequalities between the 51 American states, for the female population. The evolution of spatial inequalities in mortality seems in this case similar to the evolution observed in the French case: inequalities initially decreased and then increased. The minimum is reached in 1982 with regard to the Gini index. Thus, in the recent period, these results seem to indicate different developments with regard to international inequalities and internal inequalities.

[4] *United States Mortality DataBase*. University of California, Berkeley (USA). Available at `usa.mortality.org` (data downloaded 10/09/2018)

Figure 2.3.3: SPATIAL MORTALITY INEQUALITIES FOR WOMEN, 1980–2014



Notes: P90-100 means the share of lifespan lived by the 10% of *départements* with the highest values (compared with a uniform lifespan for all *départements)*. All inequality indicators are non-weighted by population, for women. Sample includes 90 *départements*.

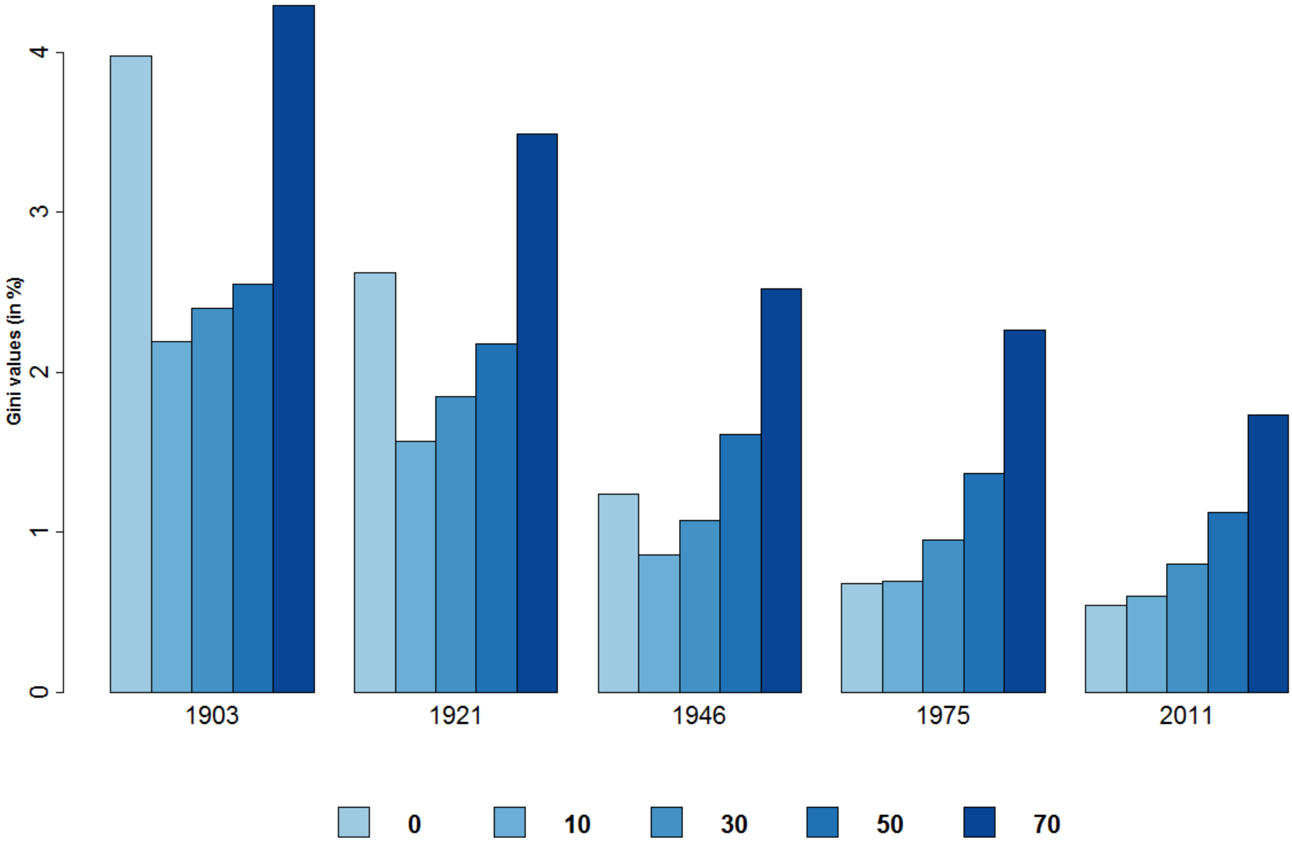
2.3.2 Role of Infant Mortality in Shrinking Spatial Inequalities

Next we analyze the role of the various age-groups in the reduction of spatial inequalities. It examines inequalities in life expectancy calculated at selected ages (0, 10, 30, 50 and 70) in 1901–2014. Figure 2.3.4 shows the Gini index for life expectancies by reference age and date.

Figure 2.3.4 shows a reduction in spatial inequalities over the century whatever the reference age. However, the reduction is greater at lower ages. The age profile of spatial inequalities changes from a U-shape to a rising curve. At present, spatial inequalities are 3 times greater for life expectancy at age 70 than at birth. Two main conclusions may be drawn from this.

First, the reduction in infant mortality played a key role in the fall in spatial inequalities. From 1901 to 2014 the reduction of the Gini index is 85% for life expectancy at birth and 65% at age 10 (the age used by Edwards and Tuljapurkar, 2005 for their international comparisons). The trend in survival rates at age 10 confirms this. Table 2.3 shows that the number of survivors per 100,000 live births rose substantially, from less than 80,000 to nearly 99,650. Not least, the inequalities between *départements* (as measured by standard deviation) shrank considerably.

Figure 2.3.4: Age profile of gini for lifespan at different ages



Notes: "10" means the Gini of lifespan at age 10. Gini is non-weighted by population, for women. Sample includes 90 *départements*.

Table 2.3: Number of survivors at age 10 for 100,000 births

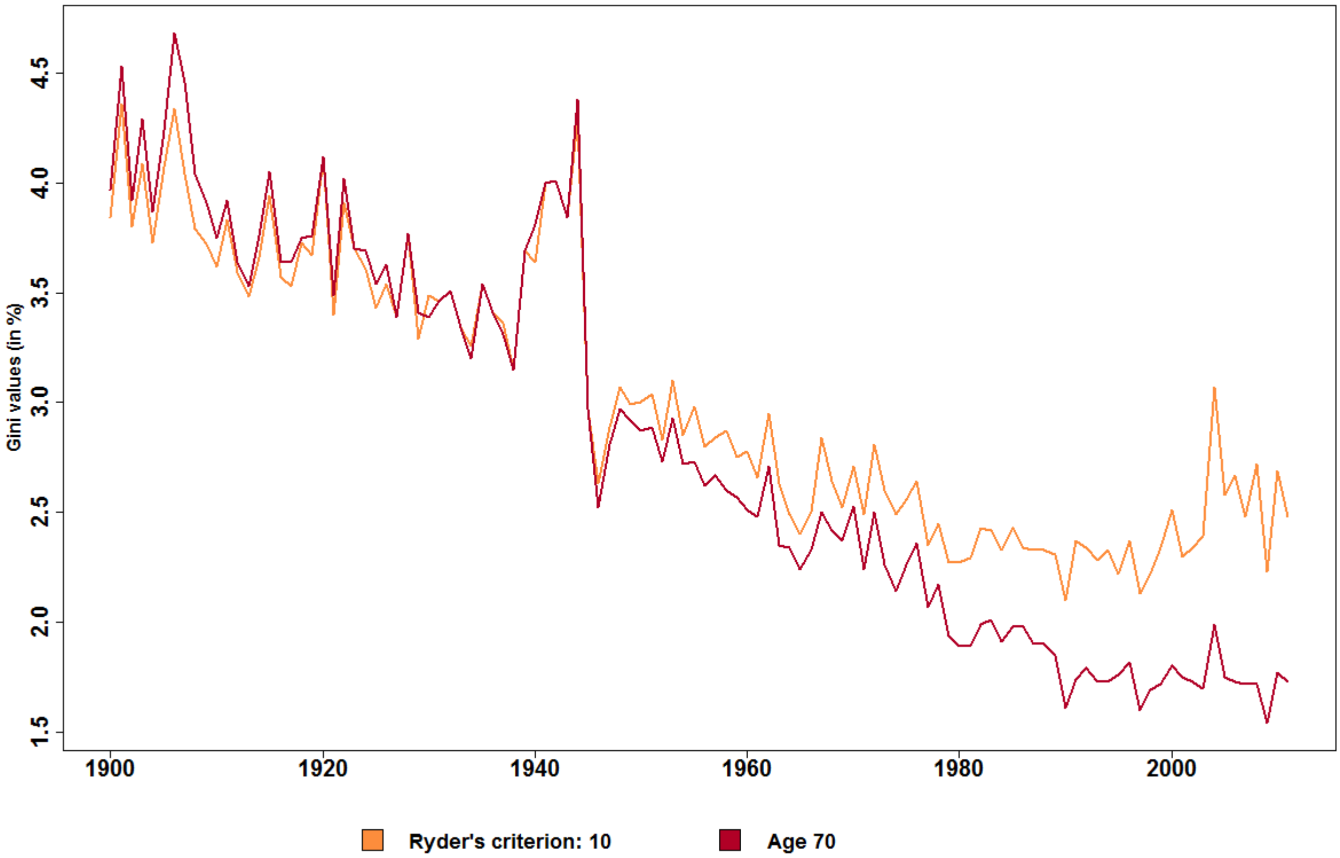
	1901	1921	1946	1975	2011
National average	79,940	84,712	90,935	98,508	99,646
Standard deviation	3,734	2,596	1,396	302	127

Notes: Sample includes 90 *départements*. Standard deviation non-weighted by population, for women.

Second, spatial inequalities at present are mainly due to differences in mortality among the oldest groups. For a given population, the epidemiological transition (Omran, 1971) causes greater variance at higher ages and lesser at lower ages (Robine, 2001). This study shows that it also affects spatial inequalities.

Interestingly, Figure 2.3.4 also shows that the trend to smaller inequalities appears to differ by reference age. Whereas the Gini index for life expectancy at birth stops falling in the 1980s, that at age 70 has fallen steadily since 1900. To put this in perspective, life expectancy at age 70 has also risen sharply during this period. It is not credible to use the same age for the start of old age when making historical comparisons (Bourdelais, 1993, d'Albis and Collard 2013). We use Ryder's criterion (1975), defining old age as beginning when someone has a life expectancy of 10 remaining years. For each year since 1900 the age was calculated at which national life expectancy was 10 years. Then the Gini index was calculated for remaining years of life in each *département* at that age. Figure 2.3.5 shows the curve of the Gini index calculated in that way and also for a set age of 70. The two curves cross in the year when national life expectancy in France was 10 years. It can be seen that the pattern is quite different and that convergence was achieved much sooner with the new Gini index. The reduction in spatial inequalities among the oldest groups stops in the early 1980s and inequalities appear to have widened again since the 1990s.

Figure 2.3.5: Gini of lifespan according to Ryder's criterion and at age 70



Notes: Ryder's criterion is the Gini of departmental lifespans at a moving age defined such as the remaining lifespan at the national level is 10 years. Age 70 means the Gini of departmental lifespans at age 70. Gini is non-weighted by population, for women. Sample includes 90 *départements*.

2.3.3 Major Changes in the Geography of French Longevity

Next we analyze variations in the *département* distribution in order to identify particular patterns in certain areas.

The data for the 90 *départements* contain a mass of information that must be systematised if the main developments are to be understood. In order to establish homogeneous geographical areas and classify these by life expectancy, cluster analysis was used. This is designed to place *départements* in classes as homogeneous as possible with regard to life expectancy during a given period. The point is to minimize variance within each class and maximize it between classes. Consequently, each class does not necessarily comprise the same number of *départements*.

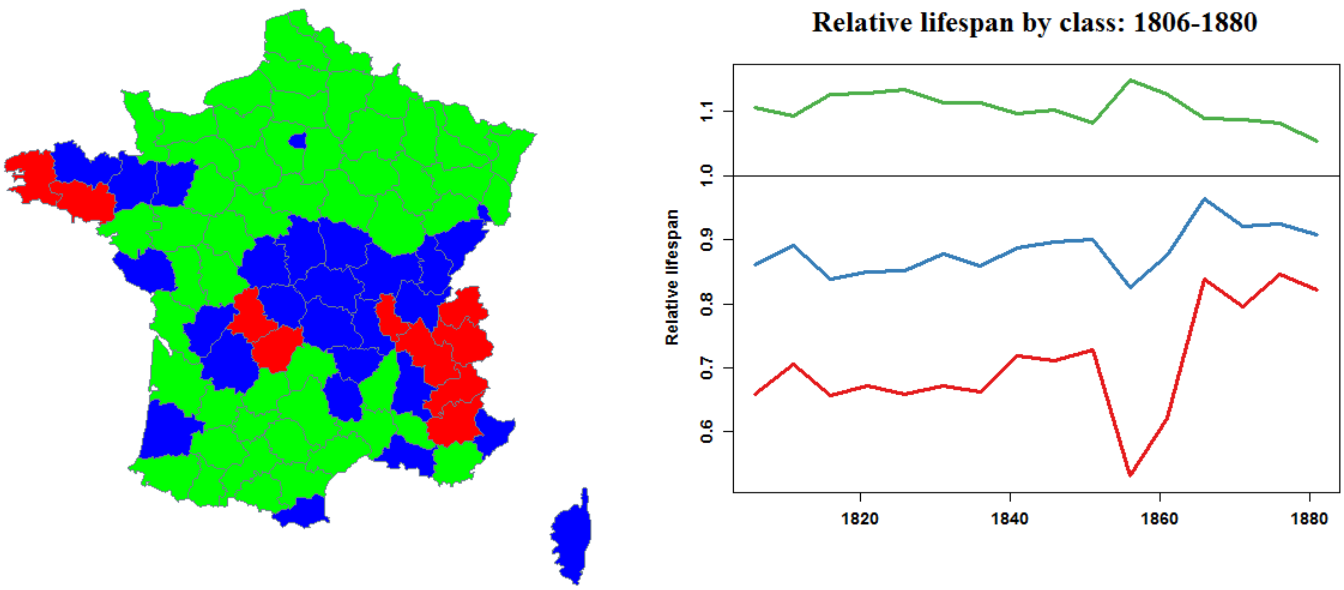
For this analysis we chose to break the data into three classes so as to identify the major trends. We applied clustering to *départements* for four sub-periods.

The first is the non-convergence phase 1806-1880. We then divided our century of convergence into two, 1881-1921 and 1922-1980. The fourth period is that of fairly settled inequalities, 1981-2014. Life expectancy data are smoothed by a moving five-year average and the years of the two World Wars are removed. To ensure that longer life expectancy does not overweight the later years, we use relative life expectancy, dividing the *département* figure by the national one for each year.

Figures 2.3.6, 2.3.7, 2.3.8 and 2.3.9 show maps of the French *départements* for each sub-period. Green *départements* are in the "top" group, where life expectancy is highest. Blue *départements* are in the "medium" group and red *départements* in the "bottom" group. The graph next to each map shows the relative life expectancy of each group over the sub-period.

The 1806-1880 period sees France cut in two by a "high mortality diagonal" from *Bretagne* to the Alps. On either side are the *départements* in the top group. This is the whole of north-east France, except for *Seine* (including Paris) and the whole of south-west France. High mortality is concentrated at the ends of the diagonal, in *Bretagne* and the Alps. The graph shows that at the start of the period the bottom group had a mean life expectancy 30% below the national mean. It also shows that inequalities between the groups began falling in 1861.

Figure 2.3.6: LIFESPAN INEQUALITY CLUSTERING, 1806–1880

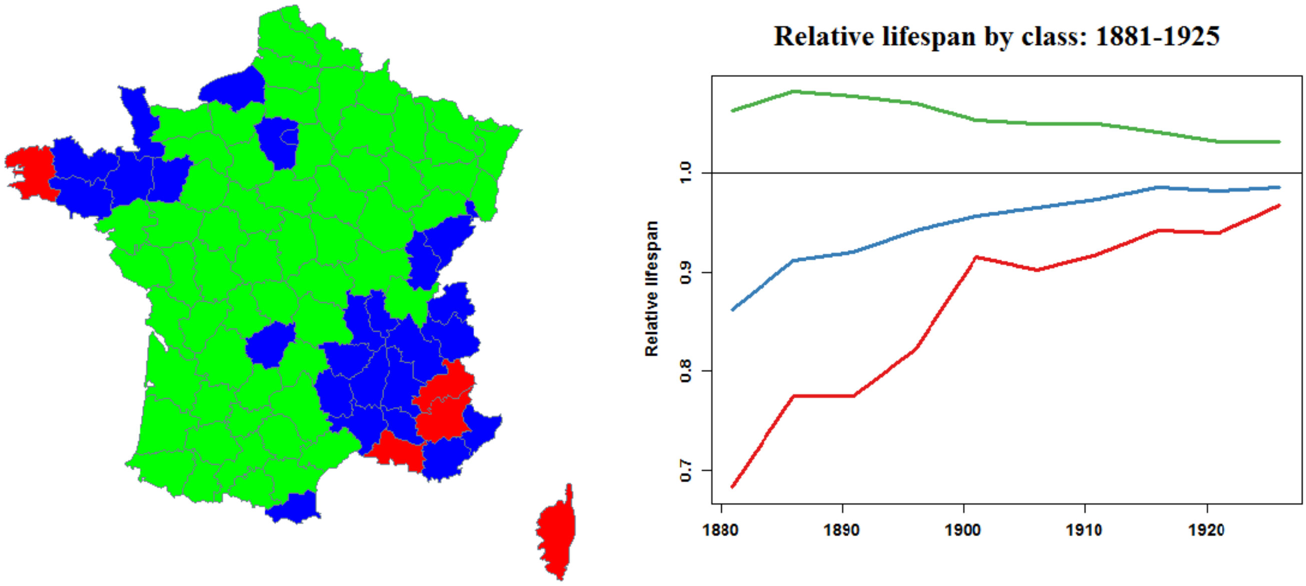


Notes: The green and red classes contain *départements* with the highest and lowest lifespan respectively. Clustering computations based on 5-year smoothed lifespan. Sample includes 90 *départements*.

The 1881-1920 period shows no great difference from the previous one, except that central France joins the top group. High mortality is still at the ends of the diagonal in *Bretagne* and the Alps. Convergence is marked, as the bottom *départements* rapidly catch up, rising from 30% below the national mean to 4% at the end of the period.

Major changes occur in 1921-1980. The whole of northern France, previously in the top group, now contains the *départements* where mortality is highest. All the *départements* along the Channel

Figure 2.3.7: LIFESPAN INEQUALITY CLUSTERING, 1881–1925



Notes: The green and red classes contain *départements* with the highest and lowest lifespan respectively. Clustering computations based on 5-year smoothed lifespan, and without war values (1914–1918). Sample includes 90 *départements*.

join *Bretagne* in the bottom group. Conversely the top group covers a wide area from central to south-west France. Note that convergence between the groups continues to progress apace.

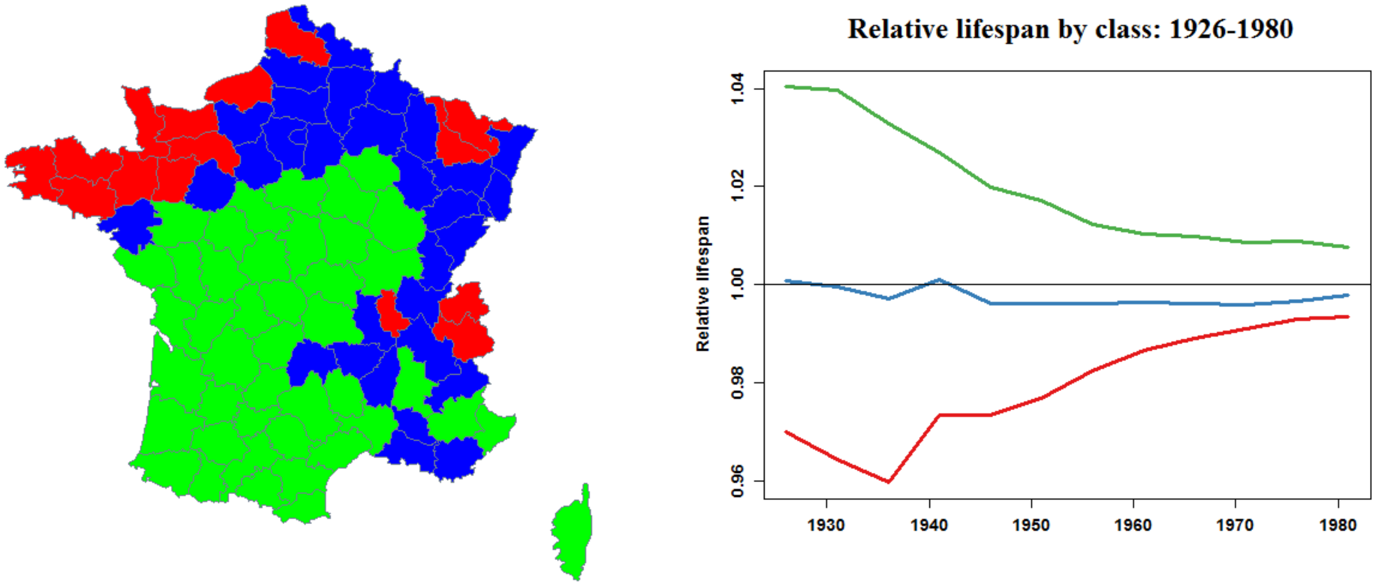
The final period, 1981-2014, shows a France once again cut in two, this time by a line from north-east to south, similar to the *diagonale du vide* ("empty diagonal") of low population density areas (Gravier, 1947, Oliveau et Doignon, 2016, Breton et al., 2017). The geography of French longevity has totally changed from what it was in the 19th century. The north of France, especially along the Belgian border, is now a high-mortality area, including *départements* like *Nord, Ardennes* and *Moselle* that once had the highest life expectancy (Fol, 2012, Laménie 2016). The only exception in the northern half of France is the *Seine département*, which has moved from the middle group to the top group. The graph for this period now shows no convergence between groups. Indeed the bottom group diverges slightly in 1995-2005. However, the gap between groups is small compared with what it was two centuries before. The bottom group's mean life expectancy is now 2% below the national mean, and the top group's does not exceed 1% above.

The main changes over the last two centuries have affected the urban *départements*, containing the major cities, and the *départements* of northern France. Their contrasting experience underlies the changing map of French mortality.

The emblematic urban *département* is *Seine*, comprising Paris. Figure 2.3.10 shows its life expectancy relative to the national mean from 1806 on. From 1856, when mortality-by-age data become more reliable, we break down this difference into age-groups (0-5, 5-20, 20-40, 40-65 and 65+). We do this by sequentially replacing *département* mortality rates for each age-group by those of the French national mean. Differences are smoothed over 5 years to improve clarity.

In the 19th century, life expectancy in Paris was well below the national mean: 18% below in

Figure 2.3.8: LIFESPAN INEQUALITY CLUSTERING, 1926–1980



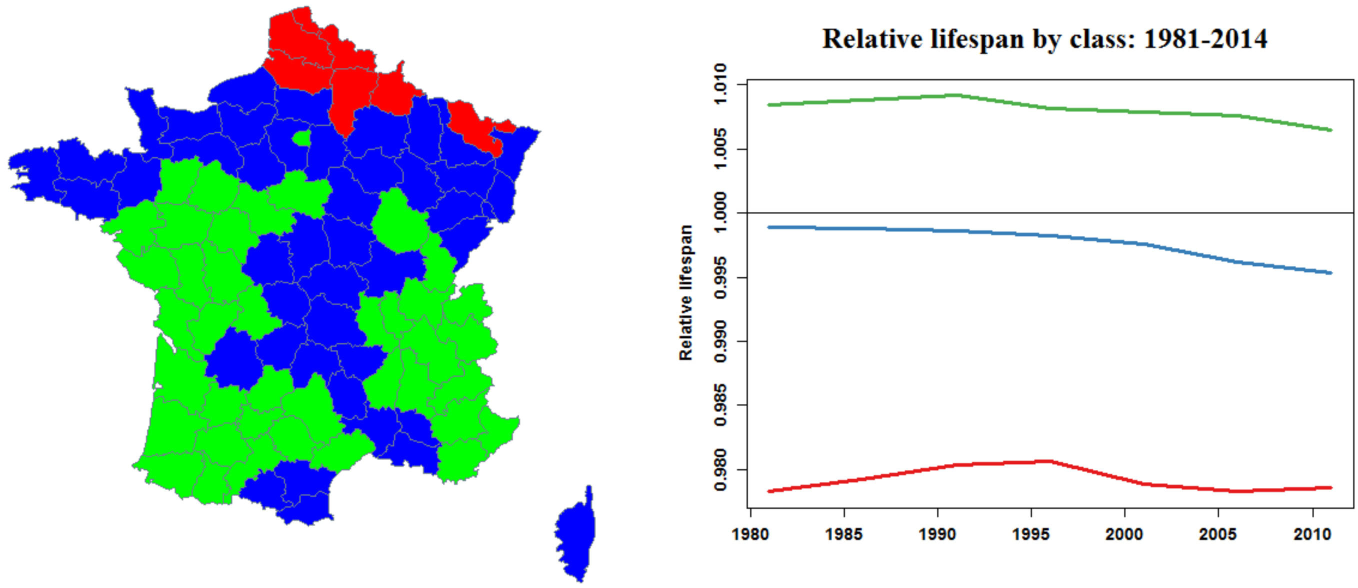
Notes: The green and red classes contain *départements* with the highest and lowest lifespan respectively. Clustering computations based on 5-year smoothed lifespan, and without war values (1939–1945). Sample includes 90 *départements*.

1816. Haines (2001) calls this the "urban penalty", which he explains mainly by the spread of infectious disease, made easier by the 2.5-fold increase in population density from 1851 to 1901 (Bonnet, 2018c). Pioneering research by Preston and Van de Walle (1978) shows that the urban penalty in the *départements* like *Seine* (Paris), *Rhône* (*Lyon*) and *Bouches-du-Rhône* (*Marseille*) was due to the quality of drinking water. This finding has been expanded by Kesztenbaum and Rosenthal (2017), who show how the gradual extension of sewer systems reduced the high mortality Paris suffered from at that time. Figure 2.3.10 confirms the link between density, mortality and poor sanitation by showing that most of this urban penalty in the 19th century comes from high mortality in the 0-5 age-group.

Seine's relative life expectancy improves throughout the 20th century: it catches up with national mortality rates and from the 1940s begins to enjoy an "urban advantage". This advantage has steadily risen since the 1990s and by 2014 was 0.84% above, equal to 10 months of life expectancy. The urban advantage of modern major cities is due to a number of factors (Vlahov et al., 2005). These cities are home to those with the highest incomes and have the best healthcare facilities (Wen et al., 2003), to the benefit of poorer people via the distribution of health expenditure (Montgomery et al., 2013). Cities also generally have more highly educated residents (Glaeser, 1999, Florida, 2002), and the gap with rural areas is increasing (Berry and Glaeser, 2005). Finally, Figure 2.3.10 shows that the urban advantage is due to lower mortality among the oldest residents, and in recent years in the 40-65 age-group.

Other *départements* that include major French cities have similar figures to Seine. *Rhône* (*Lyon*) saw its penalty become an advantage in the 1940s (cf. Appendix Figure 2.5.3). However, France's fourth largest city, *Lille*, has gone the other way. It has followed the other *départements* of northern

Figure 2.3.9: LIFESPAN INEQUALITY CLUSTERING, 1981–2014



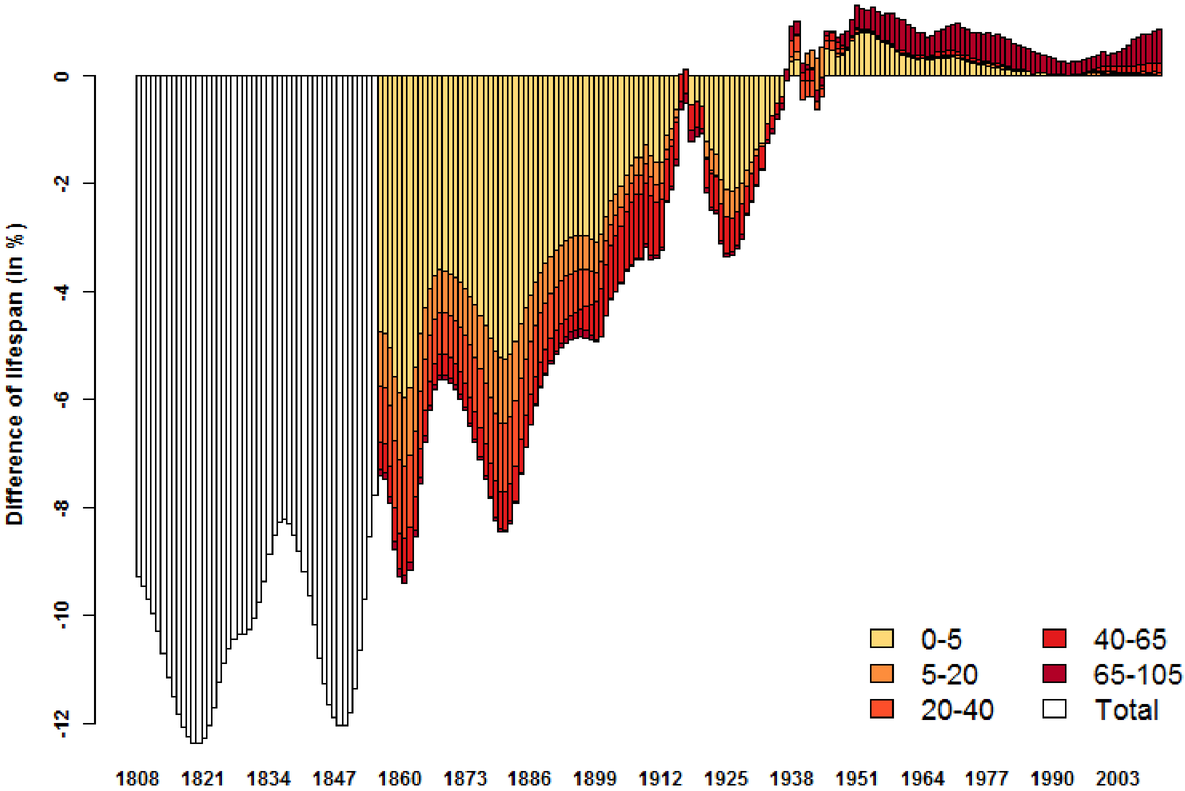
Notes: The green and red classes contain *départements* with the highest and lowest lifespan respectively. Clustering computations based on 5-year smoothed lifespan. Sample includes 90 *départements*.

and north-east France severely affected by the region's industrial decline (Zukin, 1985). Figure 2.3.11 shows relative life expectancy over time in *Nord*, which contains the city of *Lille*, and Figure 2.5.4 in the Appendix that of neighbouring *Pas-de-Calais*. *Nord* enjoyed a favourable position throughout the 19th century, except for the decade around the Franco-Prussian war, 1866-1876. This was mainly due to lower infant mortality than other regions. Starting in the 1930s, its position slipped: life expectancy in *Nord* fell below the national mean, increasingly so until the mid-1970s, when the gap was 2.5%, some 2 years. At present this gap seems to be shrinking but is still wide. It is no longer due to infant mortality, which is now insignificant in all regions, but rather to mortality after age 40. The impact of the 40-65 age-group is noticeable, accounting for 25% of the total despite extremely low mortality rates at those ages.

2.4 Conclusion

We have shown in this article that inequalities of mortality between French *départements* have considerably narrowed over the last two centuries. This trend includes a century of convergence beginning in around 1880. Only the two World Wars temporarily halted the trend. The century of convergence occurred in parallel with an increase in national mean life expectancy, in a virtuous process whereby the *départements* with the lowest life expectancy gradually caught up with the others. At present the gap between top and bottom *départements* is only 3 years 9 months, whereas it was nearly 40 years in the middle of the 19th century. However, during the last 40 years, spatial inequalities of mortality have levelled out, indeed slightly increased in 1995-2005. This is due to the worsening relative position of the bottom two deciles of *département* life expectancy. Our analysis shows that during this period, the top group of *départements* had improving mortality rates, unlike the bottom

Figure 2.3.10: Differences of lifespan between seine and france, 1806–2014



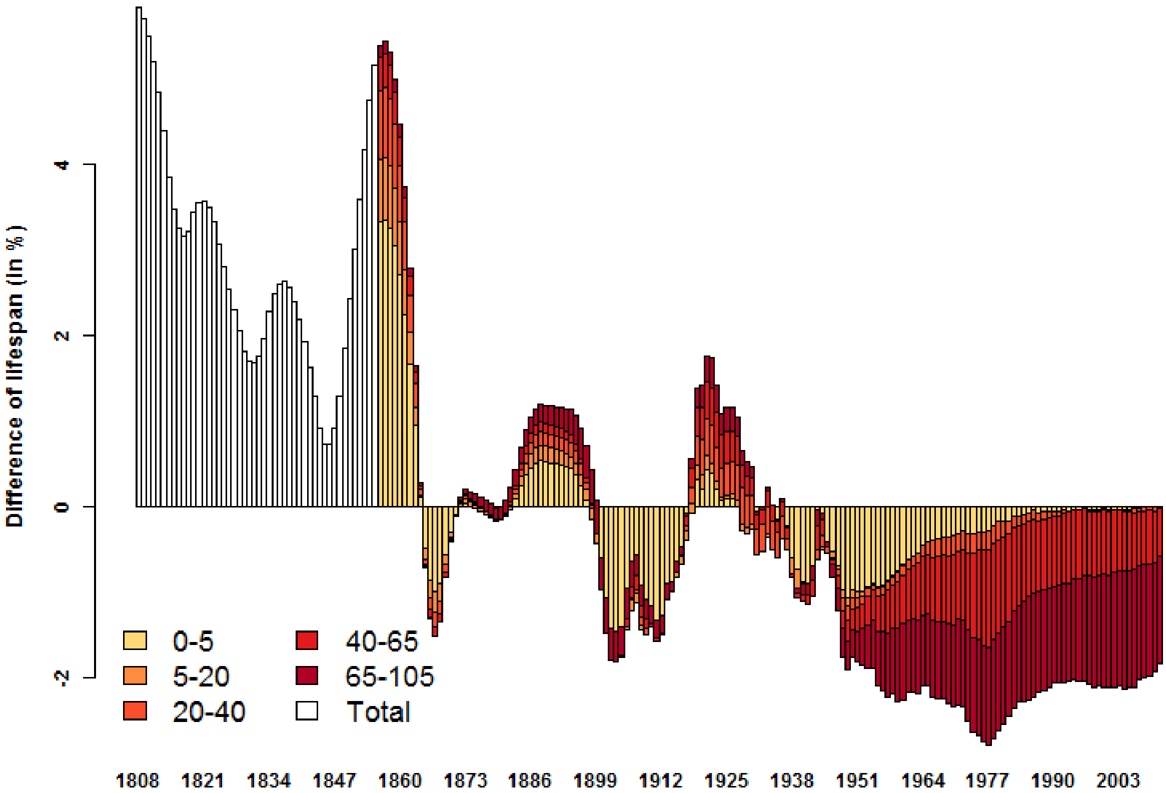
Notes: Difference of lifespan between *Seine* and France split according to the weight of each group. The split begins in 1851 since there are no reliable lifetables before this date.

group. But although spatial inequalities continue to exist and are not reducing, this is nothing like the position in the years after the French Revolution.

Our spatial analysis identifies the shifts in France's demographic map over these two centuries. Although the south-west has remained an area of low mortality and *Bretagne* one of high mortality, other regions have moved about. We have analyzed two types of *département* that have moved in opposite directions. The urban *départements*, hit by an "urban penalty" until the Second World War, now enjoy an "urban advantage", and their residents live on average 1% longer than the national mean. Conversely, the *départements* in northern France, particularly *Nord* and *Pas-de-Calais*, now have life expectancies some 2% shorter than the mean. The gap is due to higher mortality among the over-65s and the 40-65 age-group. This observation is especially remarkable because these areas were ahead of the others throughout the 19th century, with life expectancies varying from 4% to 6% above the national mean.

The demographic data at our disposal have enabled us to characterize the variations in spatial inequalities of mortality from 1806 to 2014. This is a crucial matter because inequalities of life expectancy between regions cannot be justified by public policy: there is no reason some people should die younger than others according to the *département* where they live. Health is one of an individual's set of capabilities, like income, that enable them to attain their personal goals. Public authorities are thus duty bound to seek to reduce these spatial inequalities. At present this means improving health in the *départements* of northern France. For that purpose it is essential to examine the determining factors behind spatial inequalities in life expectancy. It would be useful to correlate

Figure 2.3.11: Differences of lifespan between nord and france, 1806–2014



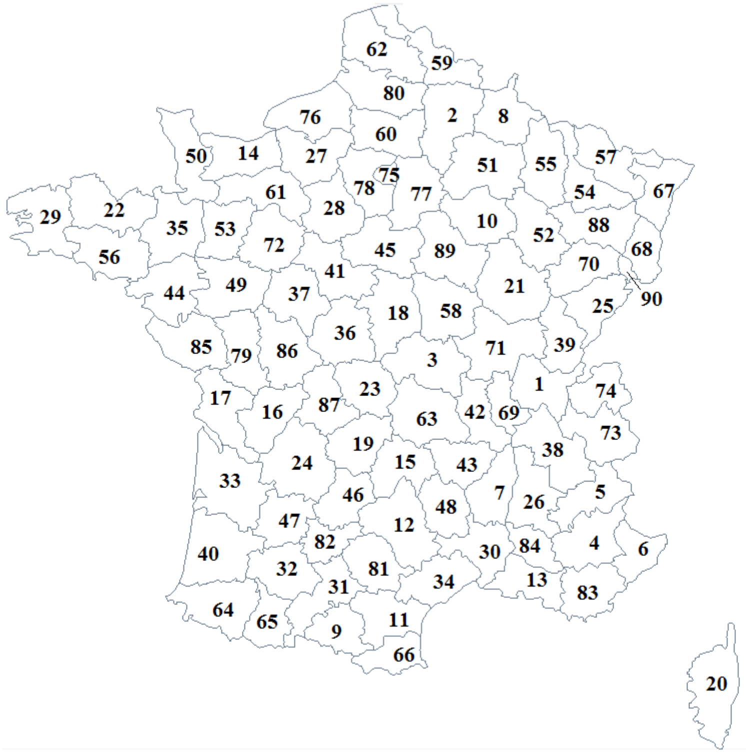
Notes: Difference of lifespan between *Nord* and France split according to the weight of each group. The split begins in 1851 since there are no reliable lifetables before this date.

our databases with epidemiological databases in order to have causes of death within the same spatial boundaries we have used. Similarly, historical socio-economic databases would be invaluable for understanding the patterns in inequalities of mortality, and thus give some perspective to recent research by Currie and Thulliez (2018).

2.5 Appendices

A1: Map of the 90 French *Départements* in 1967

Figure 2.5.1: Map of the 90 French départements in 1967



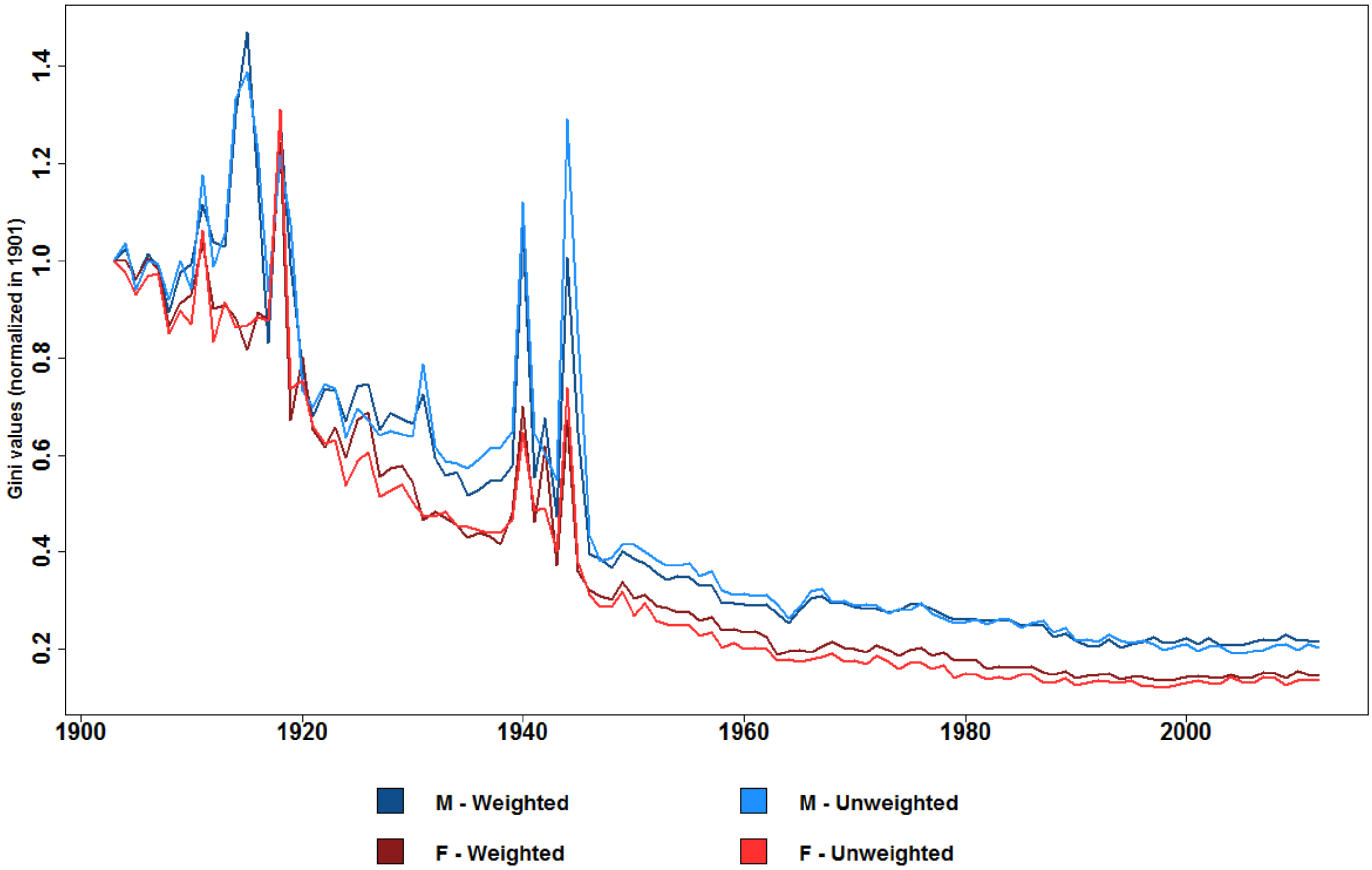
1 Ain
2 Aisne
3 Allier
4 Alpes (Basses)
5 Alpes (Hautes)
6 Alpes Maritimes
7 Ardèche
8 Ardennes
9 Ariège
10 Aube
11 Aude
12 Aveyron
13 Bouches du Rhone
14 Calvados
15 Cantal
16 Charente
17 Charente Maritime
18 Cher
19 Corrèze
20 Corse
21 Côte d'Or
22 Côtes du Nord
23 Creuse
24 Dordogne
25 Doubs
26 Drôme
27 Eure
28 Eure et Loir
29 Finistère
30 Gard
31 Garonne (Haute)
32 Gers
33 Gironde
34 Hérault
35 Ille et Vilaine
36 Indre
37 Indre et Loire
38 Isère
39 Jura
40 Landes
41 Loir et Cher
42 Loire
43 Loire (Haute)
44 Loire Inférieure
45 Loiret
46 Lot
47 Lot et Garonne
48 Lozère
49 Maine et Loire
50 Manche
51 Marne
52 Marne (Haute)
53 Mayenne
54 Meurthe et Moselle
55 Meuse
56 Morbihan
57 Moselle
58 Nièvre
59 Nord
60 Oise
61 Orne
62 Pas de Calais
63 Puy de Dome
64 Pyrénées (Basses)
65 Pyrénées (Hautes)
66 Pyrénées Orientales
67 Rhin (Bas)
68 Rhin (Haut)
69 Rhône
70 Saône (Haute)
71 Saône et Loire
72 Sarthe
73 Savoie
74 Savoie (Haute)
75 Seine
76 Seine Inférieure
77 Seine et Marne
78 Seine et Oise
79 Sèvres (Deux)
80 Somme
81 Tarn
82 Tarn et Garonne
83 Var
84 Vaucluse
85 Vendée
86 Vienne
87 Vienne (Haute)
88 Vosges
89 Yonne
90 Territoire de Belfort

Notes: Numbers used in the Bonnet (2018b)'s database. Corse is unified in this classification.

A2: Inequalities by Sex over the 1901-2014 Period, with *Départements* Weighted and Unweighted by Population

The Gini index shows a reduction in spatial inequalities of mortality since 1901 irrespective of sex or population weighting. It can be seen that inequalities have been reduced more for women than men, with a gap of some 7 points. Note too the various crises that have affected the sexes differently. The First World War temporarily increased inequalities among men but not women. Conversely, the Spanish influenza epidemic of 1919-1920 increased inequalities for both sexes, as did the Second World War. Bonnet (2018b) points out, however, that the data for the two World Wars are not entirely reliable, especially for men.

Figure 2.5.2: Gini indicator of spatial lifespan inequalities by sex and specification, 1901–2014



Notes: "M - Weighted" means Gini Indicator of spatial lifespan inequalities for men, *départements* weighted by population. "F - Weighted" is the same specification for women. Sample includes 90 *départements*.

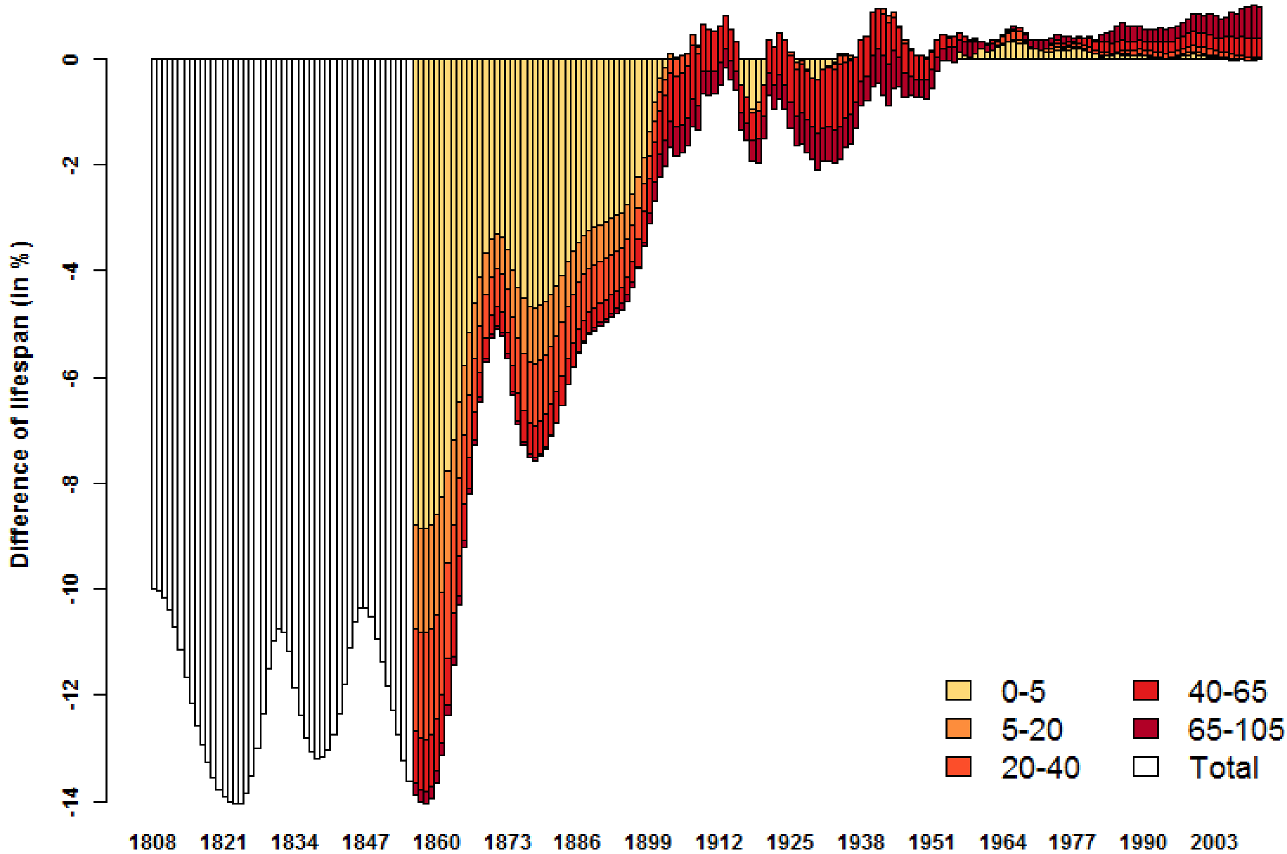
A3: Difference in Life Expectancy between Rhône and French National Mean, 1806-2014

The "urban penalty" for Rhône *département*, considerable in the 19th century – 14% in the 1850s – fell sharply to insignificant in the 1940s. Since then the *département* has enjoyed an "urban advantage" due at present to lower mortality above age 65 and to a lesser extent in the 40-65 age-group. Calculations were made for female life expectancy smoothed over 5 years. The weight of each age-group was isolated by sequentially replacing *département* mortality rates for each group by those of the national mean.

A4: Difference in Life Expectancy between Pas-de-Calais and French National Mean, 1806-2014

Until the start of the 20th century, the Pas-de-Calais *département* had a life expectancy above the French national mean. Since then it has suffered from worse mortality by some 2%. In the last 40 years this gap has levelled out and is entirely due to higher mortality among those aged 40 and

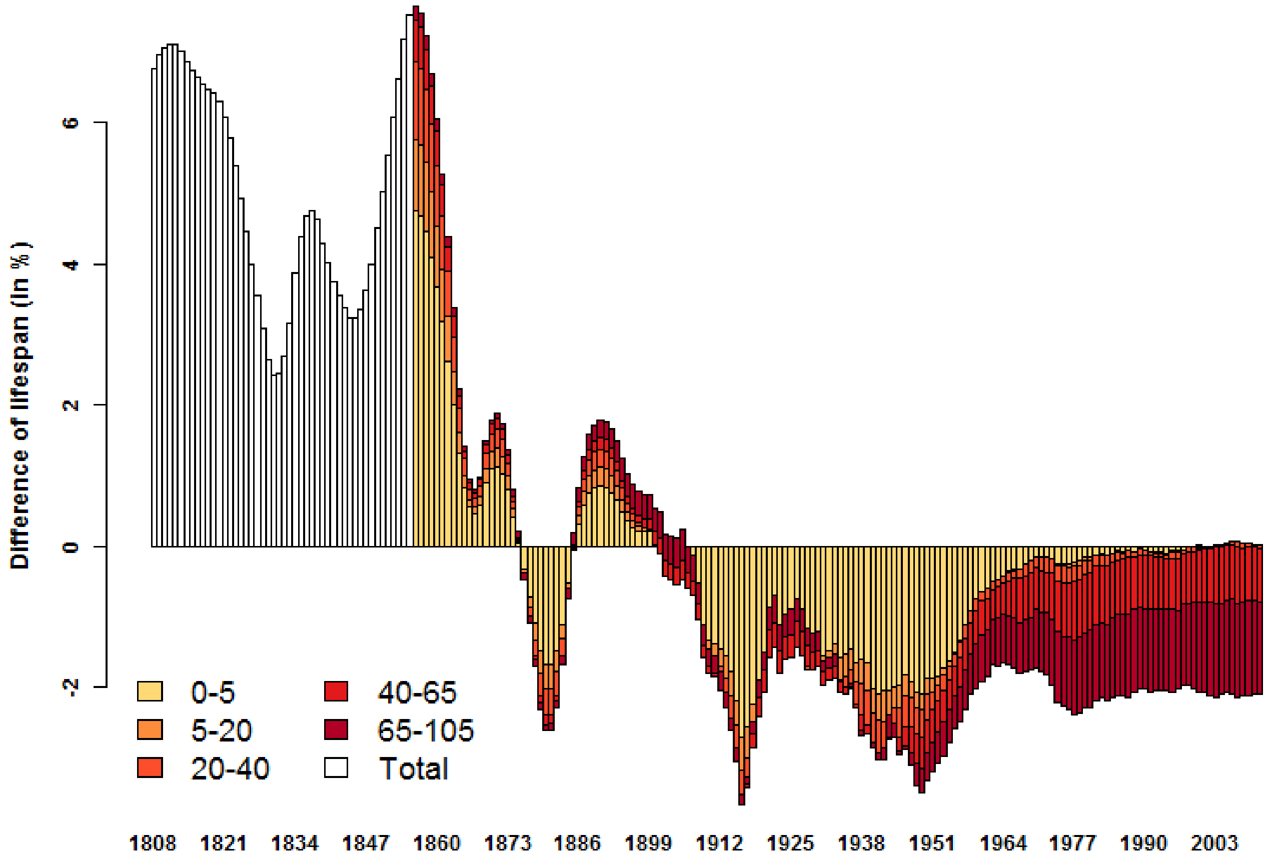
Figure 2.5.3: Differences of lifespan between rhône and france, 1806–2014



Notes: Difference of lifespan between *Rhône* and France split according to the weight of each group. The split begins in 1851 since there is no reliable lifetables before this date.

above.

Figure 2.5.4: Differences of lifespan between pas-de-calais and france, 1806–2014



Notes: Difference of lifespan between *Pas-de-Calais* and France split according to the weight of each group. The split begins in 1851 since there is no reliable lifetables before this date.

Chapter 3

Spatial Inequalities of Income and Welfare in France, 1922–2014

Abstract

In this paper, I firstly computes departmental incomes for the period 1922–2014, using demographic data (Bonnet, 2018a), economic data (Bonnet and Sotura, 2018) and tax statistics collected for the first time. Second, I use this new database to describe the evolution of spatial income inequalities per adult. I show that they have greatly decreased for nearly 100 years, and that this convergence was particularly powerful between 1950 and 1980. Moreover, I compute for the first time at local level “mortality adjusted income” by using the methodology of Fleurbaey and Gaulier (2009). I reveal that spatial inequalities have increased between 1922 and 1955, before declining. The upward phase of inequalities is explained by the gradual disappearance of the "urban penalty" of mortality. The geography of development radically changed during these last 100 years. The South-West of the country has been in a strong catch-up process, while the North-East is experiencing a deep decline.

3.1 Introduction

The end of the period of rapid growth experienced by the Western Europe and the USA as well as the intensification of globalization initiated in the 1980s has made the question of income inequalities and its evolution on long-term crucial. Piketty (2001) for France, Atkinson (2005) and Atkinson and Salverda (2005) for the United Kingdom and Netherlands, Alvaredo (2009) for Portugal for example, have shed new light on this issue. In the USA, the World Inequality Database shows that the income share of the top 10% of adults has followed an inverted U-shaped curve since the 1910s: it was 42% in 1913, 34% in 1973 at its lowest point and 47% in 2014.[1] This result can be observed in most OECD countries, with nuances regarding the recent rise in inequality. The study of inequalities from a historical and national perspective, however, has not given rise to work on spatial inequalities within countries.

In this paper, I am interested in spatial inequalities of income in France since 1922. The knowledge of spatial inequalities within a unified country over the recent period is important: these inequalities feed the question of the territorial divide, that is the resentment of populations with respect to territorial differences that should not exist. This resentment can result in extreme votes in the polls if populations believe that public policies do nothing to reduce disparities. As part of the 2017 presidential election, the North-East of France focused 9 of the 12 *départements* that voted more than 30% for the far-right party in the first round. This region is also considered one of the poorest in France because of the difficulty of reclassification of its industrial apparatus. Knowledge of the dynamics of spatial inequalities is also important. Williamson (1965) revealed an inverted U-shape curve according to the evolution of spatial inequalities: in the first time the development of the country increases these inequalities since the productive apparatus concentrates in some regions, then these inequalities should decline gradually. This result is opposed to current debates on the territorial divide, which suggest that inequality is gradually picking up again (Lessmann, 2014, Lessmann et Seidel, 2017).

The study of spatial inequalities of income in France over a long period has never been conducted. Recent literature related to spatial inequalities from a historical perspective focused on value added as the core variable, in line with Geary and Stark (2002). These researchers developed a method which need very few data in order to estimate local value added. It is based on the knowledge of sectoral value added at the national level, as well as the sectoral hourly wages and the number of individuals employed in each sector at the local level.[2] Nevertheless, the researchers did not use this method for France because the data according to the local wages were not available. Estimates of regional value added are therefore more fragmented.[3] In order to tackle the problem of missing data according to

[1] Data uploaded on May 29, 2018 on the website: `http://wid.world/en/country/etats-unis/`

[2] This method was used by Felice and Vecchi (2015) who studied Italy between 1860 and 2010, Badia-Miro et al. (2012) for Portugal between 1890 and 1980, Buyst (2010) for Belgium between 1896 and 2000, Enflo and Rosés (2015) for Sweden throughout the twentieth century or Schulze (2007) for the former Habsburg's Empire between 1870 and 1910.

[3] Three papers can however be cited in this area. The most important is Combes et al. (2011)'s one, who estimated the departmental value added for three years (1860, 1930, 2000) and studies the main trends in the evolution of spatial inequalities, both in terms of population and sectoral value added. Second, Bazot (2014) reconstructs every ten years the departmental value added between 1840 and 1911. It uses the tax statistic on the patent, an old tax

local wages, I estimate local fiscal incomes rather than value added. These estimations are based on tax statistics collected since the beginning of the income tax in 1915, at the departmental level. This provides a geographical breakdown with 90 *départements* of similar area. Moreover, the analysis of income inequalities rather than value added brings a different view: value added are studied in a productivist approach, while incomes come closer to a conception of inequality linked to welfare. Indeed, income incorporates part of the redistribution flows, particularly those related to retirement in funded systems.

In this paper, I also go further than the study of spatial inequalities of income. Nowadays, the difference between two individuals is analyzed according to their income at time t, without taking into account income over the entire life cycle. Two simple examples show that this leads to some counter-senses. Let us first imagine two agents with the same life cycle income; the first is 20-year-old and the second 40-year-old. Their income is uneven but this result is not abnormal: the oldest benefits from his past experience in the labor market. Imagine then two individuals of the same age, with the same income profile during life; the first will live 10 years more than the second. There is no inequality in the classical sense, but their situation is by definition unequal. This paper seeks to overcome the second type of shortcomings, by monetizing mortality differences between two individuals. By doing so, I am in line with Becker et al. (2005), Edwards (2012) or Fleurbaey and Gaulier (2009): they used this method to study inequalities at the international level. To my knowledge, this is the first time this method is used at the subnational level.

With the computations of "mortality adjusted income" per adult, I show how an analysis of inequalities of income per adult without taking into account mortality differences can lead to some counter-senses. The analysis of the "mortality adjusted incomes" reveals that spatial inequalities have followed an inverted U-shape – growing between 1922 and 1955 and decreasing between 1955 and 2014 – and not a monotonic decrease. This result is explained by the inverse correlation between mortality and income per adult at the beginning of the period: urban *départements* suffered from a penalty in terms of mortality, which ended after the Second World War. Geographically, "mortality adjusted incomes" per adult allows me to take a fresh look at the issue of "Shrinking Regions" (Oswalt and Rieniets, 2006, among others). I show that the South-West has been in a strong catch-up process for almost 100 years, while the North-East suffers from a significant decline.

The rest of this paper is organized as follows. In the second section I present the data and the methods that enabled me to reconstruct the departmental fiscal incomes, as well as the methodology used to monetize mortality differences. In the third section I present the results. The fourth part concludes.

3.2 Data and Methods

In this paper, I study the dynamics of the spatial inequalities of income per adult between the French *départements* since 1922. I choose this geographical unit as their administrative borders are stable since their creation in 1789. The only change which took place during the period concerns the Paris

based on wealth generated by non-agricultural activity. Third, Caruana-Galizia (2013) develops an econometric model based on sectoral employment to estimate the departmental value added between 1871 and 1911.

region. In the 1960s, *Seine* and *Seine-et-Oise* gave way to seven new *départements*: *Paris*, *Yvelines, Essonne, Hauts-de-Seine, Seine-Saint-Denis, Val-de-Marne* and *Val d'Oise*. I reconstruct the total fiscal income for the two former *départements* until 1966, and for the seven new *départements* from 1967 onwards. Moreover, I use the classification in force before the reorganization of *Ile-de-France* for the analysis of spatial inequalities, in order to get a stable geographical framework. Finally, note that my analysis focuses on the *départements* of metropolitan France. Overseas *départements* are not included because available data are too recent to carry out a long-term analysis. I use the word "national" instead of "metropolitan" to facilitate understanding of the results. My panel of *départements* comprises 90 geographical units (see map in Appendix 3.5.2).

3.2.1 Existing Databases Used

I use three sources in this paper. The first is Bonnet (2018a)'s database, which includes yearly departmental populations and age and sex-specific mortality rates between 1901 and 2014. To compute these statistics, I used population movement statistics (births and deaths by age, for each sex and each year) as well as census populations at regular intervals (by age and sex). These raw data were collected in official publications. I also retrieved raw statistics according to both military mortality during the Two World Wars and deaths in deportation. In this database, concerning the *Ile-de-France* region, all statistics are available for *Seine* and *Seine-et-Oise* until 1968, then for the seven new *départements* from 1968 onwards. In order to get the data according to the two classifications, I distribute the deaths and the populations using the proportions observed in 1968, then I recalculate the lifetables of the missing *départements*.

The second statistical source is Bonnet and Sotura (2018)'s database. In this paper we computed the departmental distributions of income for the periods 1960–1969, 1986–1998 and 2001–2014. To do so, we collected tax tabulations by *département* established by the tax authorities and available in administrative documents never officially published. Then we used the methodological protocol developed by Blanchet et al. (2017) to estimate departmental distributions of income. The main statistic I use for this paper is the income per adult of each *département*. Bonnet and Sotura (2018) calculated fiscal income according to two specifications, with and without capital gains. I choose to extract fiscal income without capital gains as I am interested in long-run developments; this specification introduces less short-run variations.

The third statistical source comes from Garbinti et al. (2016), who calculated yearly national distributions of income for the whole 20th century. This paper extends Piketty (2001)'s pioneering paper by estimating the entire income distributions thanks to Blanchet et al. (2017)'s protocol, and by calculating these distributions by adult and no longer by tax unit. Garbinti et al. (2016)'s paper contains informations on national accounts. For my study I use two of them. The first is fiscal income, which has to be distributed among the different *départements*.[4] I use the fiscal income per adult without capital gains to remain consistent with Bonnet and Sotura (2018). The second relates to annual inflation rates in order to get 2014 constant euro values.[5] I assume that the evolutions of

[4]Appendix A (Excel spreadsheet), Table A0, Column 1.
[5]Appendix A (Excel spreadsheet), Table A0, Column 17.

both departmental and national prices were identical.

3.2.2 Income per Adult between 1922 and 2014: Estimation Method

My method to estimate income per adult is based on aggregate statistics according to income tax. As Piketty (2001) recalled, the income tax was created during the First World War and has never ended up. Changes in the tax legislation have not called into question its nature: it is a progressive national tax that applies at the tax unit level (single, couple without children, couple with one or more children...). The tax is based on imposable income which gathers a large part of the income earned by adults in the tax unit. The main advantage of this tax is that it is calculated in the same way for all tax units in France, but statistical centralizations were done both at the departmental and the national level.

Bonnet and Sotura (2018) did not estimate the departmental distributions of income for the periods 1922–1959, 1970–1985 and 1999–2000 since tax tabulations were not available. Consequently, I use other raw materials to compute the fiscal income per adult of each *département*. I have collected three different aggregate statistics at the departmental level between 1922 and 2014, on an annual basis: the imposable income declared by taxable tax units, the number of tax units subject to income tax and the total of tax paid by these tax units.[6] Official publications in which the raw data have been collected are presented in Appendix 3.5.1. Note that no statistics are available for *Moselle*, *Bas-Rhin* and *Haut-Rhin* between 1939 and 1945 because these *départements* were under German administration. For these *départements*, I assume that their evolutions were synchronized with the geographically nearest *département*, namely *Vosges*. I calculate the ratios between each missing *département* and *Vosges* for each of the three variables in 1938 and 1946, and I have assumed that this ratio evolved linearly during the Second World War. Moreover, these statistics are not available for the years 1976 and 1977 as well as in 1954[7]; incomes declared by imposable tax units are not available between 1978 and 1985 and between 1999 and 2000 too.[8]

The method aims at allocating the national fiscal income used by Piketty (2001) and Garbinti et al. (2016) between the *départements* of my sample. To do so, I rely on Bonnet and Sotura (2018)'s estimates of the departmental fiscal incomes between 1960–1969 and 1986–1998. I develop an econometric model with which I estimate the values in 1922–1959, 1970–1985 and 1999–2000. Two different versions of the model are used for the whole period, to account for the availability of raw data.

In the first one, fiscal income per adult[9] relative to the national average ($F\tilde{I}^{pa}$) is estimated from

[6] All these statistics relate to tax or income collected for year X, not tax or income collected during year X. This makes a difference when income changes significantly from one year to the other.

[7] In 1954, raws statistics are available in the publications but the year events (Poujadist revolts against the tax administration) make them inconsistent.

[8] At the national level, the income tax rate (ratio of annual tax collected and total fiscal income of the year) remained stable at around 2% from 1922 to 1955, then increased steadily until 1989 – when it was 10% – before declining thereafter to reach 6% in 2014. Developments were relatively similar regarding the tax basis: 30 to 40% of fiscal income was subject to the income tax from 1922 to 1955, a percentage that increased to 60% in the 1970s before declining until the early 2000s (about 55%). A change in tax legislation suddenly increased this percentage in 2006: it reached its maximum at this date, at 70% of the total fiscal income.

[9] Adults are defined as individuals aged 20 and over.

demographic and economic variables. Demographic variables refer to the age structure of the all-sex population $\tilde{P_{ait}}$.[10] This age structure takes into account the difference of income according to the place in the life cycle: it increases from 20 to 55 years before reaching a plateau and decreasing after retirement. Globally, a population whose age structure is distorted in favor of the highest-paid age-classes should have higher fiscal income per adult than the national average.

The economic variables are imposable income per adult relative to the national average ($I\tilde{I^{pa}}$) and the share of imposable income subject to income tax at the national level ($ShII_{nat}$). Indeed, one may think that the fiscal income per adult of a *département* should be high if its imposable income per adult is itself high. However, this variable is not sufficient because the proportion of fiscal income subject to income tax changes along the period; this has an impact on the value of β. To be convinced, one can take two polar cases. In the first, the whole fiscal income is subject to income tax. Thus β equals 1: an imposable income per adult 20% higher than the national average leds to a fiscal income per adult 20% higher than the national average too. In the second, only the tax unit with the highest income is subject to income tax. Thus the share of the *département* in which this tax unit lives is 100% of imposable income, regardless of the share of fiscal income in each *département*. $ShII_{nat}$ acts as a leverage on the link between imposable income per adult and fiscal income per adult. Note that the model also includes departmental fixed effects δ_i, which take into account the specificities of each geographical unit; there is also the usual error term ε_{it}.

Finally, Equation (3.2.1) presents this first model, which is used to estimate the values between 1922 and 1959 and between 1970 and 1975:

$$F\tilde{I_{it}^{pa}} = \alpha_0 + \sum_{a=1}^{6} \alpha_a \tilde{P_{ait}} + \beta I\tilde{I_{it}^{pa}} + \gamma I\tilde{I_{it}^{pa}} \times ShII_{natt} + \delta_i + \varepsilon_{it}, \tag{3.2.1}$$

with i the *département* and t the year.

The second model takes into account that the imposable income of imposable tax units are not available in the raw materials between 1978 and 1985 and between 1999 and 2000. Like the first model (Equation (3.2.1)), it includes the departmental demography through the age structure. Moreover, I consider that the income per adult relative to the national average is high as the number of tax units per adult relative to the national average is high ($T\tilde{U^{pa}}$). The relationship goes in the same way with respect to the relative amount of tax paid by adult ($\tilde{T^{pa}}$). Equation (3.2.2) presents this model, which is used to predict the values between 1978 and 1985 and between 1999 and 2000:

$$F\tilde{I_{it}^{pa}} = \alpha_0 + \sum_{a=1}^{6} \alpha_a \tilde{P_{it}^{a}} + \beta \tilde{T_{it}^{pa}} + \gamma T\tilde{U_{it}^{pa}} + \delta_i + \varepsilon_{it}. \tag{3.2.2}$$

Table 3.1 summarizes, for each period estimated (Line 1), the support period on which the estimates are based (Line 2), the model used (Line 3) and finally the R^2 of each of the estimation periods. For the period 1922–1944, I use only the years 1960–1969 as support period because the years 1986–1998 were too far from the economic conditions at that time. In a same way, I use the years 1986–1998 to estimate values in 1999 and 2000. Table 3.1 shows that the models estimate almost exactly the income per adult relative to the national average: the R^2 is about 0.990 for each of the five subperiods.

[10] This age structure is built from seven age groups a: 0–19, 20–29, 30–39, 40–49, 50–64, 65–79 and 80 and over

The lack of imposable incomes for the period 1978–2000 has little impact on the predictive power of the model.

Table 3.1: MODEL AND SUPPORT PERIOD USED FOR EACH PERIOD WITH MISSING DATA, 1922–2014

Estimated Period	1922–1944	1945–1959	1970–1975	1978–1985	1999–2000
Support Period	1960–1969	1960–1969; 1986–1998	1960–1969; 1986–1998	1960–1969; 1986–1998	1986–1998
Model used	1	1	1	2	2
R^2	0.996	0.993	0.993	0.988	0.996

Notes: "Estimated period" is the period with missing data that I have to estimate. "Support period" is the period during which I estimate my models. "Model used" refers to the two models presented in equation (3.2.1) and (3.2.2).

In order to predict the values for the years 1954 and 1976–1977 for which the raw data are missing, I assume that the fiscal income per adult relative to the national average ($\tilde{FI}_{it}^{pa}$) evolved linearly over the period surrounding the missing years. Finally, I make a uniform adjustment so that the sum of the departmental fiscal incomes is equal to the national fiscal income collected in Garbinti et al. (2016).[11] Note that the difference between the sum of the departmental fiscal incomes and the national fiscal income is very small and about 0.5%, except between 1945 and 1959 for which the sum of departmental fiscal incomes is about 2 to 3% higher.[12]

In the rest of the paper I will use the word "income" to refer to "fiscal income without capital gains". Moreover, all values are expressed in constant euros of 2014.

3.2.3 Income and "Mortality Adjusted Income"

The analysis of inequalities between individuals or countries is usually based on distributions of value-added or income per capita that do not account for differences in mortality. This is the case for example in Williamson (1965), who showed that inter-regional inequalities followed an inverted U-shape, with an increase of inequalities in the early stages of development, followed by a gradual decline. This is also the case for studies of inequalities of income per capita, in a national context (see for example Alvaredo, 2009 for Spain, Atkinson and Salverda, 2005 for the Netherlands, and Roine and Waldenström, 2008 for Sweden) or internationally (Barro and Sala-i-Martin, 1992 for the most well-known). Yet differences in income per capita can be magnified by differences in mortality if those who earn the most are also those who live the longest. This correlation is shown for the second half of the 20th century, for example by Elo and Preston (1995), Mackenbach et al. (2003) or Waldron (2007).

[11] If Garbinti et al. (2016)'s national income was 0.5% higher than the sum of departmental incomes, I multiplied departmental incomes by 0.5%.

[12] Two other models using year-fixed effects instead of department-fixed effects have been tested. To sum up, these specifications show lower R^2 and a higher difference between the sum of departmental fiscal income and the national fiscal income. Nevertheless, it gives roughly the same results after the uniform adjustment. Since I am not interested in the values of the coefficients of the variables but in the values of fiscal incomes estimated, I have decided to present only these specifications.

Becker et al. (2005) were the first to monetize gains in life expectancy to analyze convergence at the international level. They proposed a method which allows this monetarization by using the lifecycle model with uncertain lifespan (see Yaari, 1965 and Barro and Friedman, 1977, among others). Using this method, they computed "full income" and showed that, once the differences in life expectancy gains between 1960 and 2000 are taken into account, international inequalities of “full income” per capita have evolved in a much more favorable way than inequalities of income per capita. This paper was later replicated and extended, for example, by Edwards (2012) – who introduced uncertainty about life expectancy when Becker et al. (2005) considered life expectancy as deterministic – and by Jones and Knelow (2010), who studied differences in consumption per capita rather than value added per capita.

However, as shown by d’Albis and Bonnet (2018a), the methodological framework used is fragile: one cannot calculate the "full income" in initial date, and the process of convergence is too dependent on this initial date chosen. This problem is solved using the methodology proposed by Fleurbaey and Gaulier (2009). In this paper, the authors calculate international living standards that take into account differences in income, health, working time, unemployment and inequality within countries. To compute these living standards, Fleurbaey and Gaulier (2009) define yearly the best country for each of these variables, and monetarize the difference compared to this reference country. Thus, unlike Becker et al. (2005), the initial date problem no longer arises.

I use this methodology for the first time to analyze the spatial inequalities of living standards within a country. With the data available I calculate the “mortality adjusted income” of each *département*: this income takes into account inter-departmental differences in life expectancy. Each *département* is analyzed through a representative individual: its income is equal to the departmental average income, and its life expectancy is equal to the departmental mean life. For each year, a reference *department* is used to monetarize differences in life expectancy. Section 3.5.3 presents formally the method used to calculate the adjusted income of mortality. It is in line with the works of d’Albis and Bonnet (2018a).

3.2.4 Analysis of Spatial Income Inequalities

To analyze the evolution of spatial income inequalities (for both "classical incomes" and “mortality adjusted incomes”), I use differences of departmental incomes per adult. I implicitly assume that there are no intra-departmental inequalities. I weight each *département* by the share of their population in the national population. I choose these weights as the welfare of the population is central in this study, which requires not giving too much weight to the depopulated *départements*. On the other hand, equal-weights could be explained by a public policy point of view where the territory prevails.

In order not to leave aside the question of spatial planning, I focus on the spatial distribution of income on the national territory too. Indeed, as Ayuda (2010) explained it, the decline in inequality can be paralleled by an increasingly uneven distribution of this income due to inter-departmental migrations. If the richest *départements* are also the most densely populated, the migration of the population from the poorest *départements* to the richest ones increases the unequal distribution of income on the national landscape. This creates other problems, namely in terms of desertification or lower political weight in specific territories. Consequently, I use the departmental densities of income, defined as the ratio between the total income and the number of km^2 in a *département*. I take into account the number of km^2 in each *département* since the geographical units are not homogeneous. For example, *Gironde* in the South-West has an area of 10,375 km^2 when that of its neighbor *Tarn-et-Garonne* is only 3,718 km^2. This transformation makes it possible to bring the *départements* back to the same standard.

For the analysis of inequalities, there are a large number of indicators that can be used (see Cowell, 2011 for a review). Among them, one can differentiate indicators that take into account the whole distribution of income (Gini index or Theil index are the more popular) or the indicators that take into account only a part of the distribution, including the rank (difference between the lowest value and the highest value) and the interquantile interval (the difference or ratio between the value defining the most favored x % and the most disadvantaged x %). Recent literature, supported particularly by Piketty (2001) and by researchers at Wid.World, highlighted why an analysis of indicators targeting specific parts of the income distribution is relevant. Piketty (2001) analyzes the shares of the 0.1%, 1%, 10% of tax units with the highest incomes, but also the share of the 50% with the lowest incomes. These indicators are independent of average values, in the same way as the Gini or Theil indices.

In this paper, I use the Gini index to get an idea of the global evolution of inequalities. Nevertheless, this indicator suffers from some weaknesses: a drop in the Gini index can be explained both by a redistribution of income from the first decile to the second or a redistribution of income from the ninth decile to the tenth. Following Piketty (2001), I overcome this problem by calculating the shares of the 10% of adults with the highest incomes in the national income (the first decile, called P90–100), but also the shares of each of the other deciles. In this way, I split the departmental distribution to analyze each part that composes it. To deduce the final inequality indicators used to analyze inequalities of income per adult, I relate these shares to what they would be in the case of perfect equality. Thus, if the top-decile share is equal to 15% (compared to 10% in the case of an equal distribution), the inequality indicator is equal to 50%. Semantically, this means that these individuals have an income 50% higher than the average income. This transformation homogenizes the values of all the inequality indicators. Note that I use the same indicators to analyze the spatial distribution of income: I calculate the share of the 10% of km^2 whose income densities are the highest, and so on.

3.3 Results

3.3.1 The Three Phases in the Decrease of Spatial Inequalities of Income per Adult

I begin by analyzing the spatial inequalities of income per adult for the period 1922–2014. To do so, I use the Gini index and the shares of different fractiles throughout the distribution. I split the distribution into seven different parts: the first millime (P99.9–100), the first decile except the first millime (P90–99.9), the second decile (P80–90), the third and fourth decile (P60–80) and so on until the last two deciles (P0–20). These indicators are calculated by weighting the *départements* by their share in the adult population of both sexes. To limit erratic short-term variations, indicators are smoothed over 5 years. The points for 1924 are therefore the average values observed between 1922 and 1926. Since I have collected the total income and the total tax for each *département*, I represent these indicators for both before-tax and after-tax income.

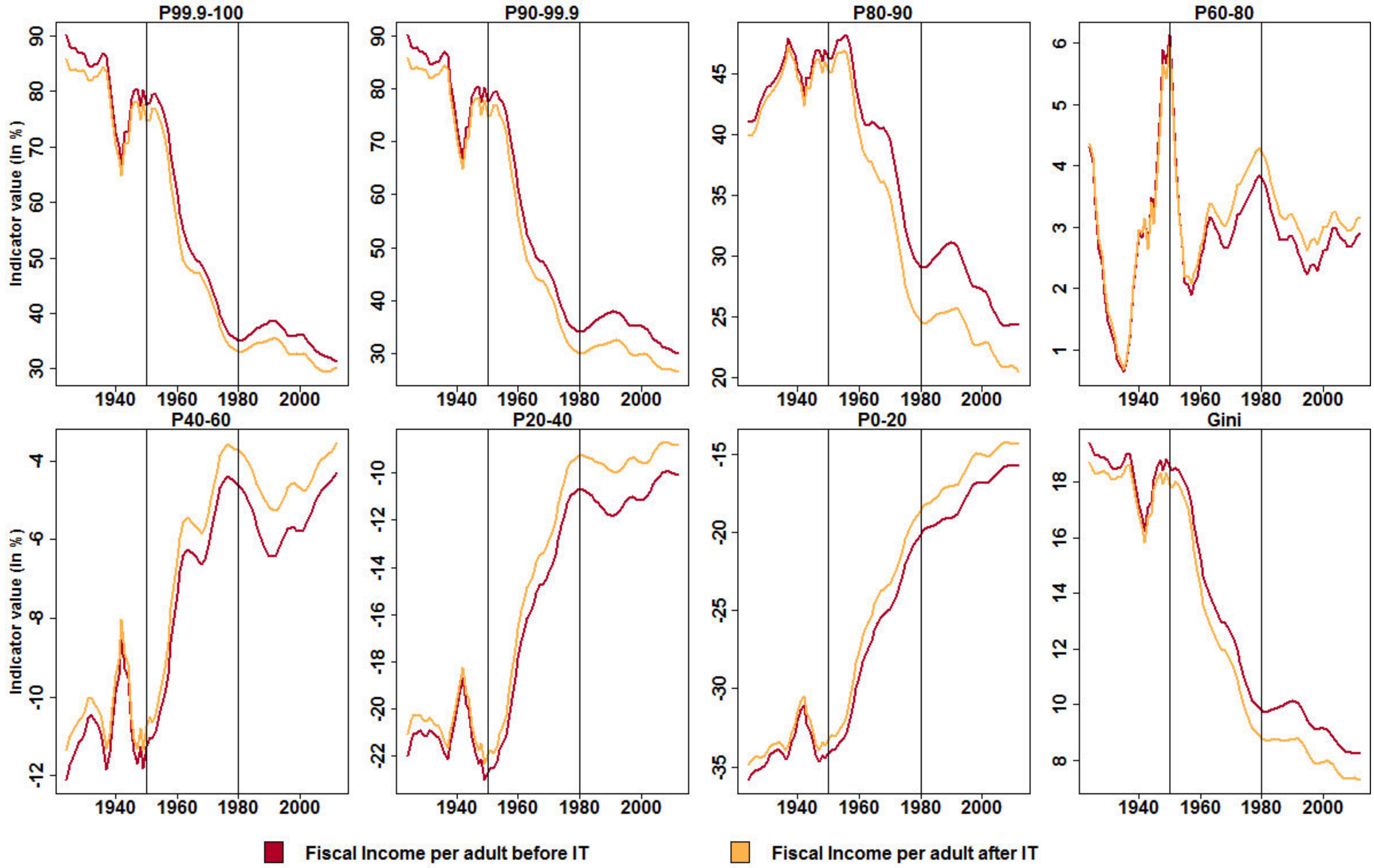
At first I focus on the spatial inequalities of income per adult before tax. Figure 3.3.1 shows that they have fallen since 1922: the Gini index has dropped from 0.194 to 0.082, a decrease of nearly 60%. In 1922, the average income per adult of the first decile was 90% above the national average, compared to 35% under this average according to the last two deciles. These figures were in 2014 30% and -15%, respectively. Beyond this overview, the analysis of the Gini index helps to sequence the process in 3 phases.

From 1922 to 1950, inequality remained almost stable: the Gini index decreased by only 8% to 0.185. During this period, one can see that this stagnation hides 3 ambivalent phenomena: the decrease of the first decile share (P90–99.9 and P99.9–100) and the increase of the last two deciles share (P0–20), which pushed down inequalities, and the increase of the second decile share (P80–90) which pushed them upwards. In other words, the relative advantage of the richest territories – namely *Seine*, *Seine-et-Oise* and *Rhône* – decreases in favor of the a-little-less richest and the poorest territories. This very slight decline in inequalities went hand in hand with a stagnation (or almost) of the average income per adult, from €3,670 in 1922 to €4,215 in 1950 (an annual growth rate of 0.5%). On the other hand, one can note that the Second World War temporarily reduced these inequalities. This phenomenon is largely explained by the disorganization or destruction of the industrial and tertiary productive apparatus in the urban *départements*, which gave more weight to agricultural activities in the national income. These activities were mainly located in the poorest territories (the transitory increase of the last eight deciles share proves this statement).

A second phase began in 1950 and ended in 1980. During this period, the Gini index goes from 0.185 to 0.097, a fall of nearly 50%. One can note that the catch-up process affects the last six deciles, to the detriment of the first two. As such, the average income per adult of the last two deciles, which was 34% below the national average in 1950, is only 20% lower in 1980. In contrast, the average income per adult in the first decile was 77% higher than the national average in 1950, compared to only 34% in 1980. It is interesting to note that this period is almost perfectly juxtaposed with the "glorious 30" (Fourastié, 1979): between 1950 and 1980, the average income per adult was multiplied by a 3.7 factor (an annual growth rate of 4.7%). For these reasons, I call this phase "glorious 30 of spatial inequalities".

Finally, from 1980 to 2014, inequalities followed an inverted U-shaped curve: they increased from 1980 to 1990 (the Gini index gains 0.004 point to 0.101), then decreased from 1990 onwards. For the whole period, one can see that the convergence process is mainly explained by the decline in the share of the second decile (which lose 5 points) and the rise of the last two deciles (which gained 5 points). Conversely, the first decile as well as the third, fourth, fifth and sixth deciles have no longer participated in the process of convergence. It would seem, then, that middle-income territories reached a glass ceiling – with an income per adult about 7 to 8 percent below the national average – that the poorest deciles have not yet reached. This slight decrease in inequality have been paralleled by a structural change in average income growth: it increased by 43% in 34 years (an annual growth rate of 1.1%).

Figure 3.3.1: SPATIAL INEQUALITIES OF INCOME PER ADULT, 1922-2014



Notes: P99.9-100 means the share of income earned by the 0.1% of adults with the highest incomes (compared with a distribution in which any adult earns the same income*)*. IT means Income Tax. All inequality indicators are weighted by adult population, for both sexes. Inequality indicators are smoothed over a 5-year period. Sample includes 90 *départements*.

The collection of the yearly tax paid in each *département*, necessary to estimate the fiscal income per adult, makes it possible to compute inequalities both before and after the income tax. The evolution of income tax legislation has been very chaotic since its creation, as explained by Piketty (2001, Chapter 4). I summarize his story in a few words. This tax was created in 1915 during the First World War. Originally, it consisted of a progressive tax on all incomes, and proportional taxes specific to each category of income (*"cédules de revenus"*). This persisted until the end of the Second World War, when it was replaced by both a proportional and a progressive tax on total income. It then took the form that we know today, that is a single progressive tax called *"Impôt sur le Revenu des Personnes Physiques"*. For this study, I have collected only the data on the progressive tax that persisted through the decades; I am not interested in other forms of direct taxation.

Even if the income tax is not intended to redistribute income between *départements*, one can see on Figure 3.3.1 that it reduced inequalities through a higher taxation of *départements* with the highest incomes compared to the *départements* with the lowest incomes. In 2014, the average income per adult for the last two deciles (P0–20) was 15.8% and 14.4% below the national average according to pre-tax income and post-tax income, a difference of 9%. These results are found for all fractiles. However, the strongest impact on inequalities achieved for intermediate deciles: the second decile inequality indicator is 16% lower after taking into account the income tax (17% lower according to the fifth and sixth deciles).

Concerning the dynamics, the income tax reduced the Gini index by only 2 to 3% between 1920 and 1950. Subsequently, this percentage increased rapidly – 14% at its maximum at the end of the 20th century – then decreased to reach about 11% today. This evolution follows faithfully the one of the overall tax rate (ratio of the tax collected during the year to the fiscal income, at the national level). To conclude, the income tax magnified the decline in inequality: the Gini index dropped by 59% in terms of income per adult before tax, and by 63% according to income per adult after tax.

3.3.2 The Virtuous Convergence of Income per Adult during the Second Half of the 20^{th} Century

I continue my analysis by studying the evolution of both spatial inequalities of income per adult and spatial inequalities of income densities. Section 3.3.1 has revealed that the departmental incomes per adult converged over the period 1922–2014, with two temporary breaks during the Second World War and between 1980 and 1990. This does not mean however that income is more evenly spread over the national territory. To answer this question, it is necessary to study the evolution of inequalities of income densities. By analyzing in parallel the inequalities of incomes per adult and densities of income per km^2, one can present two possible scenarios.

In the first, which I call "Unbalanced Convergence", income per adult converges while income per km^2 diverges. In this scenario, migration plays an important role: adults from the most disadvantaged territories (which are also the least densely populated) migrate to the most favored territories, reinforcing imbalances of the spatial distribution of population and income. This scenario creates financial problems, especially if the tax system is highly decentralized. Indeed, if local resources are correlated with local incomes, the gradual disappearance of the tax basis gradually reduces these resources. Since some public expenditures are proportional to the territories area and not to the population living there, this scarcity of resources leads to financial shortcomings. In the second scenario, which I call "Virtuous Convergence", income per adult and densities of income per km^2 converge. In this case, the catching up of the most disadvantaged *départements* is endogenous since it is not explained by migrations but by a development of the territory.

To see what were the scenarios between 1922 and 2014 in France, I calculate the variations in the Gini index of incomes per adult and incomes per km^2 between each five-year period from 1922 to 2014 (using pre-tax income). As an example, the point 1924 is the couple of Gini variations between 1922–1926 and 1927–1931. The points I get for each period are placed on a quadrant diagram (Figure 3.3.2): in the South-East quadrant there are "Unbalanced Convergence" phases. In addition, the evolution of inequality indicators for income densities per km^2, in line with those of Figure 3.3.1, are presented in Appendix 3.5.4.

Figure 3.3.2 recalls that inequalities of incomes per adult has continually declined between 1922 and 2014: only the 1944 point is not in the lower two quadrants. As seen in Section 3.3.1, this is explained by the rebound of inequalities following the sharp decline observed during the Second World War. By taking into account the evolution of inequalities according to income density per km^2, I shed light to 3 phases in the convergence process.

The first extends from 1922 to 1953 (in Red on the diagram): during this period, spatial inequalities of income per adult fell in parallel with an increase in spatial inequalities of income per km^2. Indeed, apart from the points associated with the Second World War (1934 and 1939), these points are located in the two eastern quadrants. The Gini's evolution (Appendix 3.5.4) confirms this result: it increases by 7% between 1922 and 1953 to reach 0.555. This period is characterized by a process of "Unbalanced Convergence". This is in line with Bonnet (2018b)'s results, who showed that the spatial distribution of the French population had been increasingly uneven during the first half of the 20th century; this process is called "hyper-centralization thwarted". This period is the end of the rural exodus, during which a large part of the population first migrated exclusively to the Paris region and also to the second largest cities.

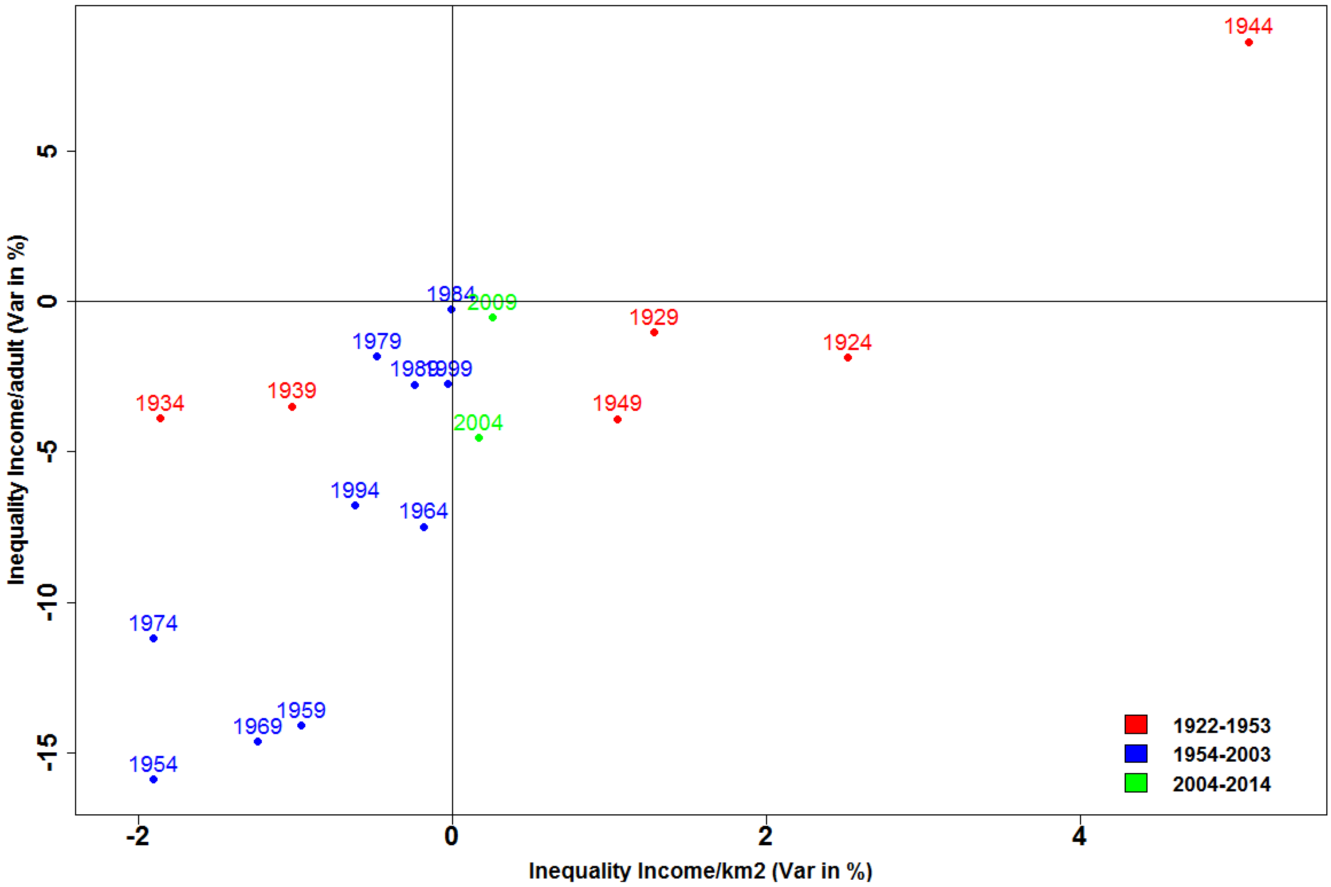
From 1954 to 2003 there was a half-century of "Virtuous Convergence" (in Blue on the diagram). One can see that all the points are in the South-West quadrant. Thus, the second half of the 20th century both reduced the spatial inequalities of income per adult and distributed in a more balanced way this income over the national landscape. Two sub-periods are visible during this half-century. The first extends from 1954 to 1978 and coincides with what I have called the "glorious 30 of spatial inequalities" (1954 to 1974 points on the diagram). One can see that the two convergences are strong, with five-year rates about 12% according to the decrease of spatial inequalities of income per adult and 1-2% according to spatial inequalities of income per km^2. On the other hand, the period 1979–2003 shows much lower rates of variation, 4% and 0-1%, respectively.

Finally, the beginning of the 21st century reveals a new phase of "Unbalanced Convergence". Indeed, the 2004 and 2009 points are in the South-East quadrant. Moreover, note that the 2009 point presents almost a stagnation of the spatial inequalities of income per adult. Appendix 3.5.4 provides a better understanding of this trend reversal. Firstly this is due to a decrease in the convergence process of the first decile, with an upward curve of the P90–99.9 fractile share in recent years. It is also explained by a reversal in the share of intermediate deciles (P20–40 and P40–60), which drove convergence during the second half of the 20th century.

3.3.3 The Evolution of Spatial Inequalities of "Mortality Adjusted Income" per Adult

My analysis now focuses on the differences of spatial inequalities of income per adult, depending on whether spatial differences in mortality are taken into account or not. To do so, I calculate "mortality adjusted incomes" using the methodology presented in Section 3.2.3. Table 3.2 shows a summary of distributions of departmental differences between income (defined as usual) and "mortality adjusted income" (both per adult, in % of income). By construction, the minimum equals 0 for the *département* with the highest life expectancy (Line 1). One can see that they became smaller: the maximum

Figure 3.3.2: SPATIAL CONVERGENCE OF INCOME PER ADULT: VIRTUOUS OR UNBALANCED?



Notes: Variation of inequality of Income per adult for 1934 is the variation of Gini Index between the period 1932–1936 and the 1937–1941 one. I use the triennal period 2012–2014 to compute values for 2009. Gini Index for Income per adult is weighted by the population of adult, and Gini Index for Income per km^2 is weighted by km^2. All computations are done with fiscal income before income tax. Sample includes 90 *départements*.

difference is 26.9% in 1926, compared to 4.4% in 2013 (Line 5). The results are similar for the median: values are 11.6% and 1.6%, respectively (Line 3). These results are consistent with d'Albis and Bonnet (2018b), which showed that spatial inequalities in mortality decreased significantly between 1880 and 1980, before stagnating over the recent period. This stagnation appears in Table 3.2: between 1999 and 2013, the maximum difference increased by 0.2 point to 4.4%.

With these results, I compute the inequality indicators according to "mortality adjusted income" per adult and compare them to income per adult ones. Figure 3.3.3 revealed these indicators for both specifications. Overall, taking into account the inter-departmental differences in mortality reduces inequalities in the old period but increases them over the recent period: the Gini index according to incomes per adult (Red Curve) was 0.194 in 1922, while it was only 0.165 for the "mortality adjusted incomes" per adult (Blue Curve). These figures were 0.082 and 0.085 in 2014.

Concerning the dynamics, evolutions are not the same: while inequality of incomes per adult has decreased since 1922, inequalities of "mortality adjusted incomes" per adult first increased between 1922 and 1955 (the Gini index went from 0.165 to 0.182), before declining significantly thereafter. To explain this process, one can see that the most developed *départements* (P80–90 and P90–100) suffered a mortality penalty in years 1920–1940. Conversely, the least developed *départements* ben-

Table 3.2: Departmental distributions of differences between income and mortality adjusted income per adult

	1926	1936	1946	1962	1982	1999	2013
Min.	0	0	0	0	0	0	0
25%	-8.8	-3.1	-4.2	-2.1	-1.3	-1.3	-1.1
Median	-11.6	-5.5	-6	-3	-1.9	-1.8	-1.6
75%	-14.5	-7.6	-7.6	-4.2	-2.5	-2.5	-1.9
Max.	-26.9	-12	-13.9	-6.5	-5.2	-4.2	-4.4

Notes: Differences are in %. "Mortality Adjusted Incomes" per adult are calculated following the methodology exposed in Section 3.2.3. "25%" means that the 25% of the departmental differences between Income and "Mortality Adjusted Income" per adult are upper this level. Sample includes 90 *départements*.

efited from higher life expectancies. There was an inverse correlation between income per adult and mortality. These results support those of Preston and Van de Walle (1979), Kesztenbaum and Rosenthal (2017) or d'Albis and Bonnet (2018b) concerning the "urban penalty". For example, d'Albis and Bonnet (2018b) showed that *Seine* (in which Paris is located) suffered a "urban penalty" until the end of the Second World War. It is also in this *département* that income per adult was the most important at that time. The convergence of mortality conditions between the urban *départements* (namely *Seine*, *Seine-et-Oise* and *Rhône)* and the rest of the country between 1920 and 1940 increased the relative "mortality adjusted income" of these *départements*. As a result, this increased inequalities during the period. This process was not offset by lower income inequalities as it remained almost stable at that time.

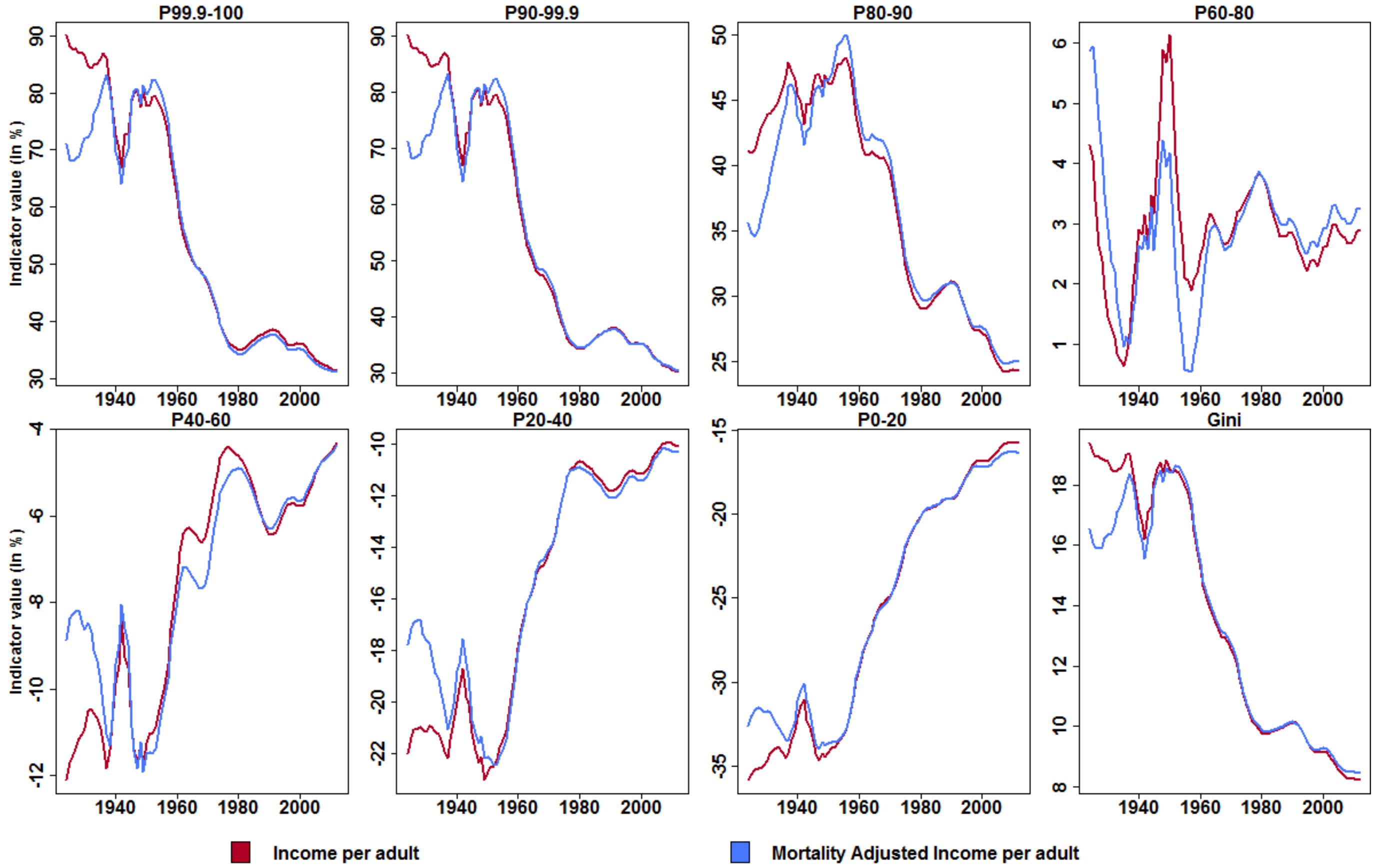
In recent times, the differences between the curves are quite small. This is explained by two phenomena. The first is the spatial convergence of mortality revealed by d'Albis and Bonnet (2018b): the gap in life expectancy at birth was 13 years old in 1922 compared to 3 years and 9 months in 2014. The second is the method, which monetize differences of mortality in the computations of "mortality adjusted incomes". From the point of view of a newly born agent, this method discounts future income streams and gives very little weight to differences in mortality at older ages. Nowadays, it is at these ages that spatial differences are most visible: infant mortality, high in the 1920s and very unevenly distributed, has almost disappeared throughout France. Finally, note that the correlation between income and health is positive today. It can be seen for all the deciles except the first: the shares of the total income held by the second (P80–90), third and fourth deciles (P60–80) increase when the differences in mortality are taken into account, whereas they decrease for the seventh, eighth (P20–40), ninth and tenth deciles (P0–20).

3.3.4 Changes in the Geography of Development in France

Finally, I am interested in the Geography of develoment and its evolution between 1922 and 2014. To do so, I calculate the departmental "mortality adjusted incomes" per adult relative to the national average for these two years. Figure 3.3.4 and Appendix 3.5.5 present the results.

In 1922, one can see that the *départements* with the highest incomes were in the Paris region and

Figure 3.3.3: SPATIAL INEQUALITIES OF INCOME AND MORTALITY ADJUSTED INCOME PER ADULT, 1922–2014

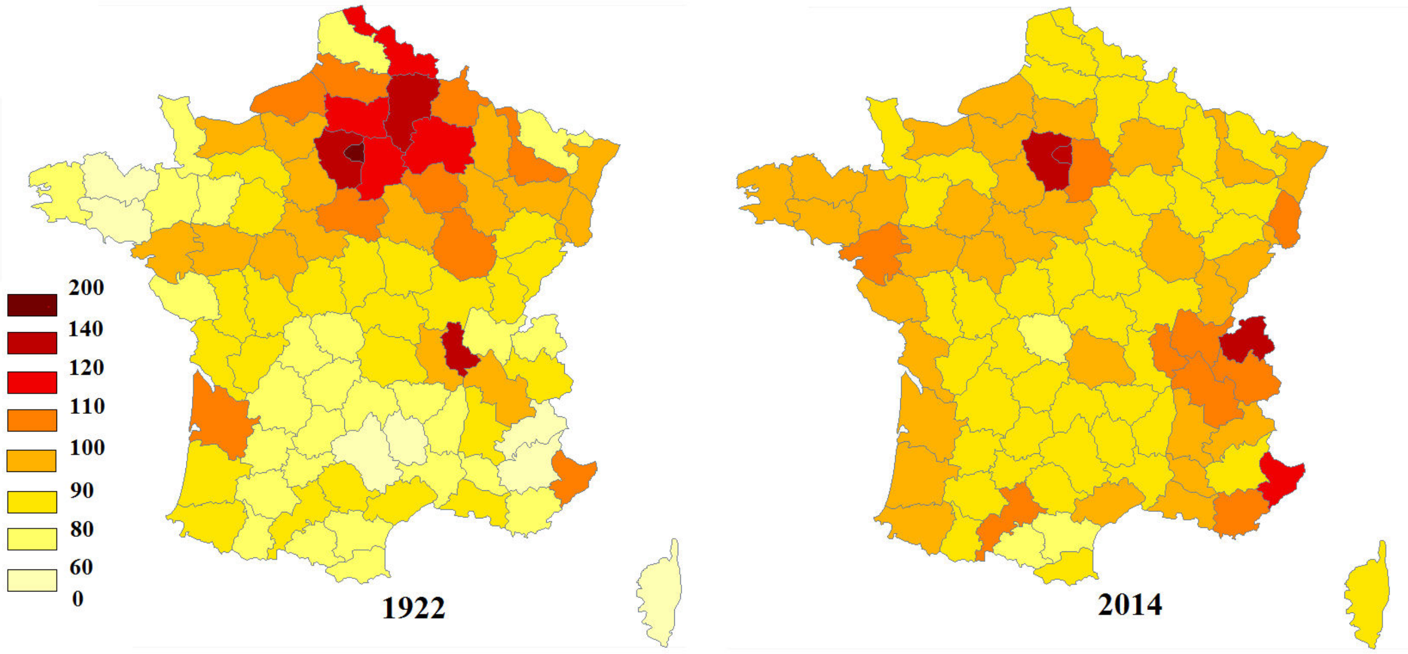


Notes: P99.9-100 means the share of fiscal income earned by the 0.1% of adults with the highest incomes (compared with a distribution in which any adult earns the same income). "Mortality Adjusted Incomes" are calculated following the methodology exposed in Section 3.2.3. All inequality indicators are weighted by adult population, for both sexes. Inequality indicators are smoothed over a 5-year period. Sample includes 90 *départements*.

the North-East of the country. For example, the "mortality adjusted income" in % of the national average reaches 181% in *Seine*, 132% in *Seine-et-Oise*, 120% in *Aisne* or 113% in *Nord*. *Pas-de-Calais* and *Moselle* were an exception with an income per adult 25% lower than the national average. On the other hand, all the *départements* that were lagging behind (values 90% or lower) were located south of a line linking *Manche* and *Haute-Savoie*. Within this geographical zone, *Bretagne*, the South of the *Massif Central* (*Aveyron* and *Lozère*) as well as the *départements* close to the Alps (*Basses-Alpes, Hautes-Alpes*) and *Corse* had incomes 50% lower than the national average. Some *départements* were doing better, namely *Alpes-Maritimes*, the *Lyon* region (*Rhône, Isère, Loire*), *Gironde* and the lower *Loire* Valley (*Indre-et-Loire*, *Maine-et-Loire*, *Loire-Inférieure*).

In 2014, the French geography of development has changed significantly. The Paris region keeps a significant advantage, with an income per adult equals to 130% of the national average. The relative position has however decreased sharply in *Seine* in just a century. The Eastern borders (Germany, Switzerland, Italy) are privileged: *Haute-Savoie*, for example, enjoys a high income per adult as well as the country's highest life expectancy. As a result, the "mortality adjusted income" is 30% higher than the national average. The Atlantic Arc, from *Côtes-du-Nord* to *Pyrénées-Atlantique*, emerges as a zone of relative prosperity: between 1922 and 2014, the income per adult converged towards the national average, or exceeded it in *Loire-Inférieure* (which hosts *Nantes*). Finally, the geographical area in which *départements* have an income 10% or more below the national average links Belgium and Spain borders. This area of relative underdevelopment, which is frequently called the "empty diagonal" (see Oliveau and Doignon, 2016 for example), has therefore changed since 1922. Moreover, one can see the effects of the catch up process during the mid-20th century, which benefited to the most backward *départements*: in 2014, there is no *département* with a "mortality adjusted income" lower than 25% of the national average (the minimum is reached in *Creuse* where it equals 75.5%).

Figure 3.3.4: Relative mortality adjusted income per adult, 1922 and 2014



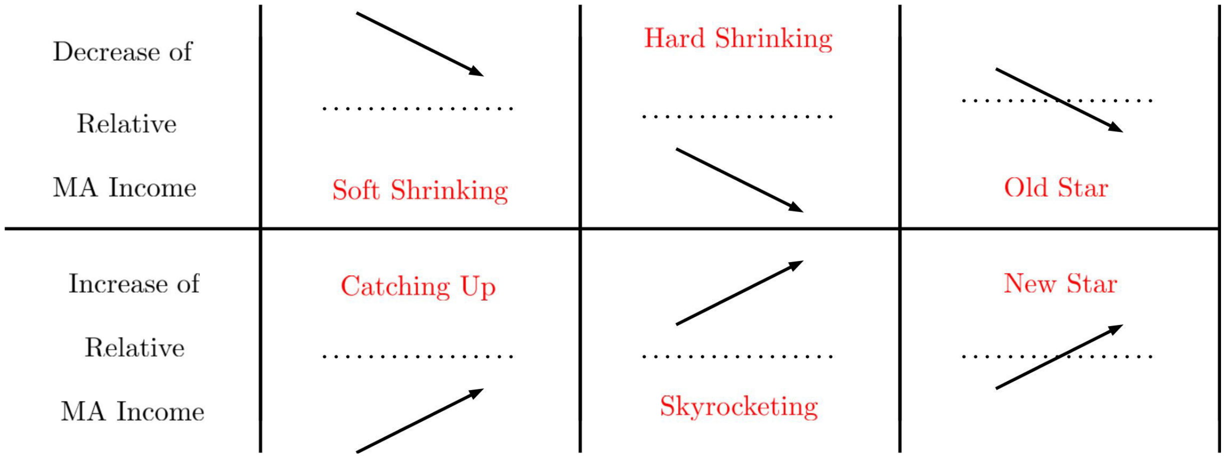
Notes: Values are in % of the national average. "Mortality Adjusted Incomes" per adult are calculated following the methodology exposed in Section 3.2.3. Sample includes 90 *départements*.

Using the variation of relative "mortality adjusted incomes" per adult between 1922 and 2014, I classify *départements* into six categories presented in Figure 3.3.5. This classification highlights regions in relative decline: this comes close to what the literature names "Shrinking Cities" or "Shrinking Regions" (see for example Oswalt and Reniets, 2006, Turok and Mykhnenko, 2007, Fol, 2012, Bontje et al, 2012). It differs however since shrinking cities or regions are those whose population decreases between two dates. In this paper, I analyze the variation of the income per adult and not the variation of the population. Moreover, the literature focuses on regions whose population decreases in absolute value while my analysis focuses on changes in relative value. This choice is explained by two reasons. The first is that the "mortality adjusted incomes" per adult globally increased sharply between 1922 and 2014 because of both the increase in income during the "30 glorious" and the convergence of mortality conditions. The second is that a territory may feel downgraded even though its income per adult has increased during the period.

According to Figure 3.3.5, the declining *départements* are classified in three categories: the *départements* whose "mortality adjusted incomes" per adult at the beginning and the end of the period are higher than the national average ("Soft Shrinking"), those for which they are lower ("Hard

Shrinking”) and finally those which fall below the national average during the period (“Old Star”). Conversely, among the *départements* whose relative situation has improved, one can find those for which income per adult at the beginning and the end of the period are lower than the national average (“Catching Up”), those for which they are higher (“Skyrocketing”) and those who have risen above the national average during the period (“New Star”).

Figure 3.3.5: Departmental classification according to the variation of relative mortality adjusted income per adult



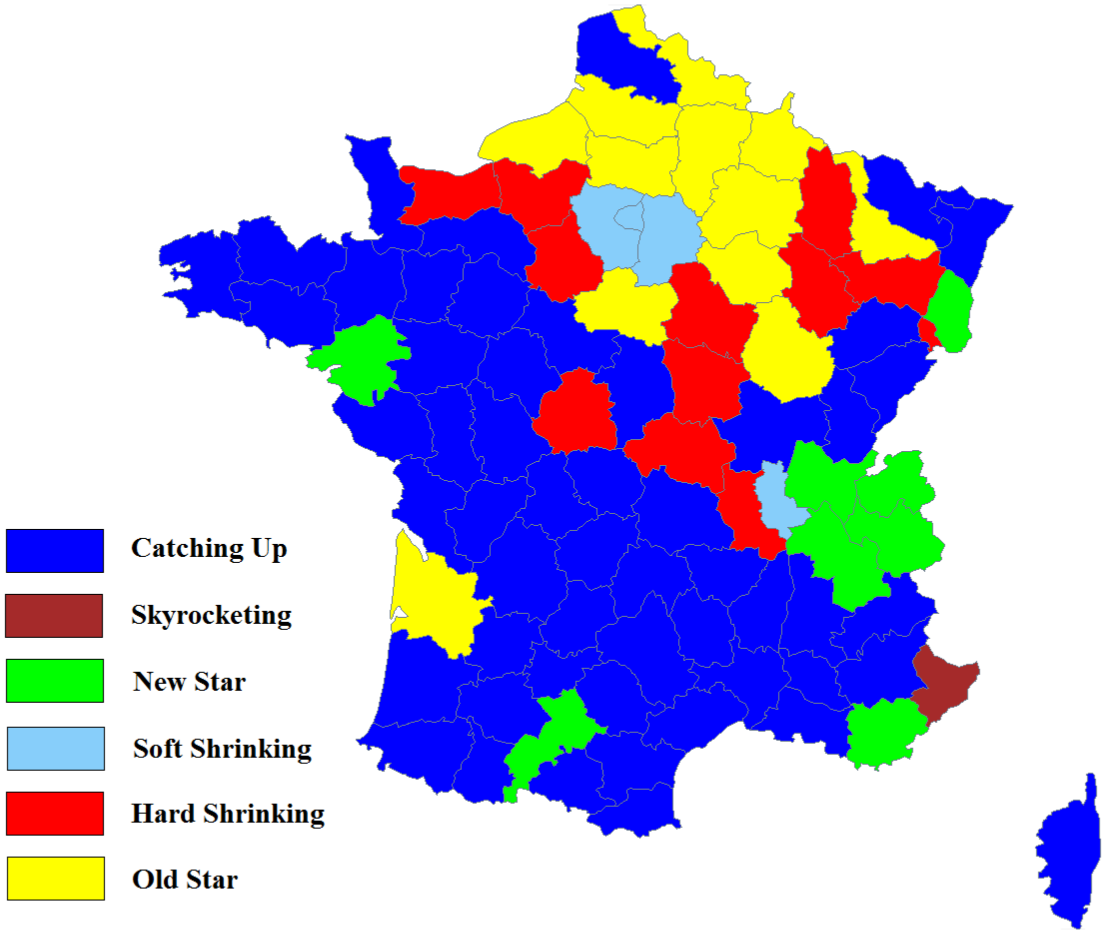
Notes: MA means “Mortality Adjusted”. Departmental classification built according to the variation of “Mortality Adjusted Incomes” per adult. “Mortality Adjusted Incomes” per adult are calculated following the methodology exposed in Section 3.2.3.

Figure 3.3.6 applies this classification to the French *départements* for the period 1922–2014. Figure 3.3.7 represents the evolution of “mortality adjusted income” per adult for 6 *départements* corresponding to the six categories of the classification. Since I am interested in long-term developments, the values have been smoothed by 5-year period to limit erratic evolutions.

A large part of the South-West, beyond the fictitious line connecting *Manche* to *Haute-Savoie*, has benefited from a catch-up process during the period. 53 *départements* are in the "Catching Up" category, more than half of them. *Lozère* is emblematic. One can see in Figure 3.3.7 that the “mortality adjusted income” per adult was 50% lower than the national average in 1922; this gap is less than 20% today. Within this geographical area, I distinguish the "New Stars", i.e. *départements* close to the Swiss border (*Ain, Isère, Savoie, Haute-Savoie*) as well as *Var*, *Loire-Inférieure* and *Haute-Garonne*. These *départements* have not only caught up, they are now enjoying “mortality adjusted incomes” per adult above the national average. For example, value in *Haute-Garonne* was 20% lower in 1922, but 3% higher nowadays. This reversal is even more important in *Haute-Savoie* where the figures are -20% and +30%, respectively. Finally, the advantage in *Alpes-Maritimes* has increased a little over time. As Figure 3.3.7 reveals, this evolution was not monotonous since this *département* was lagging behind during the late 1950s.

North-East of the line connecting *Manche* to *Haute-Savoie*, *départements* are almost all in relative decline. They include primarily the Paris region (*Seine, Seine-et-Oise, Seine-et-Marne*): with *Rhône,* they are the only *départements* whose relative positions declines but remains above the national

Figure 3.3.6: Departmental classification, 1922–2014



Notes: Departmental classification according to the variation of relative "Mortality Adjusted Incomes" per adult between 1922 and 2014. "Mortality Adjusted Incomes" per adult are calculated following the methodology exposed in Section 3.2.3. Sample includes 90 *départements*.

average. This decline is particularly strong for *Seine* (decrease of 50 points to 130% in 2014) and *Rhône*. For the latter, only the Second World War temporarily stopped the process; in recent years, relative "mortality adjusted income" has stabilized around 110%. This area hosts the "Old Stars" too, which in 1922 were the most developed *départements* but are now behind. For these *départements,* the relative decline is the most difficult since they enjoyed relative opulence during the beginning of the 20th century. It includes *Picardie* (*Aisne*, *Oise*, *Somme*), *Seine-Maritime*, *Nord*, *Loiret*, *Meurthe-et-Moselle* and those belonging to a line connecting *Ardennes* to *Côte d'Or*. Overall, relative "mortality adjusted incomes" per adult decreased by 20 points (19 points in *Somme* at 85% in 2014, 27 points in *Nord* at 86%). The case of *Ardennes* shows the timing of this decline, particularly strong before the Second World War. These results highlight the analysis conducted by Laménie (2016) for this *département*. Lastly, the relative "mortality adjusted incomes" of some *départements* decreased, whereas they were already lower than the national average in 1922. They are located in three areas around *Nièvre*, *Eure* and *Vosges*. The decline is the most sensitive in the latter one (14 points in *Meuse*, 10 points in *Haute-Marne* and 7 points in *Vosges*).

Finally, one can note the contribution of the "mortality adjusted income" specification rather than usual income. Appendix 3.5.6 reveals the same map as Figure 3.3.6 built using incomes per

Figure 3.3.7: Departmental evolutions of relative mortality adjusted income per adult, 1922–2014

Notes: "MAI" means "Mortality Adjusted Incomes", which are calculated following the methodology exposed in Section 3.2.3. Values are smoothed over a 5-year period.

adult. One can see that the number of *départements* in catching up process is higher (61 against 53). In addition, the south of the Paris region seems to be part of the catch-up zone, which is not the case in the specification I have adopted. Indeed, this territory enjoyed a significant advantage regarding mortality, which is no longer true nowadays. Finally, the category of *Ardennes* switches ("Hard Shrinking" by considering income per adult, "Old Star" with my specification). This is also explained by mortality differentials since this *département* had one of the highest life expectancies at the beginning of the 20th century.

3.4 Conclusion

In this paper, I have first proposed a methodology to estimate the departmental fiscal incomes concerning the periods 1922–1959 and 1970–1985. These computations are based on a demographic database (Bonnet, 2018a), an economic database (Bonnet and Sotura, 2018) as well as statistics on the income tax (imposable income declared, total tax collected, number of imposable tax units) used for the first time. These statistics have been introduced in an econometric model calibrated on periods 1960–1969 and 1986–1998 during which departmental fiscal incomes are well known.

I have then used departmental fiscal incomes of the period 1922–2014 to study the evolution of spatial inequalities of income per adults in France. I have shown that these inequalities have decreased over the period. This decline has followed a three-phases process. Between 1950 and 1980, the convergence was particularly strong; the rapid decline of the richest *départements* (*Seine*, *Rhône*, *Seine-et-Oise*) as well as the catching up of the least favored *départements* has allowed global convergence. In the recent period, spatial inequalities are falling much more weakly as the advantage of the most developed *départements* is no longer diminishing and the convergence of the intermediate *départements* is ending up. The analysis of both spatial inequalities of income per adult and spatial inequalities of income per km^2 shows that the years 1950–1980 could be described as the "glorious 30 of spatial inequalities". Indeed, this period saw both the rapid decline in spatial inequalities of income per adult but also a much more balanced spatial distribution of income.

In a second step, I have shown that the analysis of “mortality adjusted incomes” per adult reveals some phenomena hidden by the analysis of income per adult. To compute these “mortality adjusted incomes”, I have used the methodology of Fleurbaey and Gaulier (2009), which monetize the differences in life expectancy between a *département* and the *département* with the highest life expectancy. These calculations are repeated for each year, in order to get yearly “mortality adjusted incomes” by *département*. Using this specification, I have shown that spatial inequalities of income per adult have not declined continuously since 1922 but have followed an inverted U-shape curve: the maximum of this curve was reached at the end of the Second World War. This difference is mainly due to the fact that urban *départements* such as *Seine* and *Rhône* have for a long time undergone a "urban penalty" regarding mortality. These *départements* were also those in which incomes per adult were the highest. The convergence of mortality conditions between urban and rural *départements* therefore initially increased inequalities, before a reversal in the period 1950–1980 due to the decline of income inequalities.

The analysis of the geography of development and its evolvements has shown that some relative positions have changed over time. While much of the South-West of the country is in a strong catching up process, the North-East is in decline. This area is divided into two categories. A first in which the *départements* which benefited from an income per adult higher than the national average are laggind behind today *(Picardie, Nord, Ardennes ...*), and a second in which *départements* were late in 1922 and whose relative position has declined during the period (*Meuse, Haute-Marne, Vosges* as examples). I have demonstrated that taking into account differences in life expectancy in the computations of “mortality adjusted incomes” per adult changed the classification obtained with an analysis of usual incomes per adult. As such, the southern Paris region, which benefited from advantageous mortality conditions at the beginning of the 20th century, is not in a catch-up phase but rather in a phase of relative decline.

With these new data, I have presented the evolution of spatial inequalities of income and “mortality adjusted income” between 1922 and 2014. This question is all the more important as these spatial inequalities can hardly be explained from a public policy point of view. In a unified country like France, it is difficult to justify higher incomes in some regions. The only acceptable justification that can be made relates to the demographic and economic structures. With respect to the demographic structure, the theory of life-cycle-income tells us that income increases with age, before decreasing at retirement. Thus, a *département* where the share of young adults and retirees is above the national average should have a lower income without calling for a specific public policy. Similarly, with equal characteristics, the wages of men and women are not the same, although this phenomenon is diminishing. Taking into account the gender distribution within departmental populations could therefore help to explain the differences. With respect to the economic structure, the most profitable jobs are concentrated in the big cities: the urban *départements* therefore have a natural advantage, which could explain inter-departmental differences. For all these reasons, the collection of data by age, sex and production structure at the departmental level would allow to go further the description of the facts, and shed light to their potential explanations.

In addition, this paper has sketched an analysis of the tax system impact on spatial inequalities: beyond the fact that income tax increases the relative income of the most lagging *départements*, the gap between inequalities before and after income tax has increased over time. Nevertheless, the income tax is not the only one to make spatial inequalities more acceptable: this is also true concerning social security contributions, wealth-based taxes or value-added-based taxes for example. Comprehensive knowledge of the evolution of the socio-fiscal system impact on spatial inequalities is on the agenda; it can be considered as crucial nowadays, as the issue of territorial divide is more and more popular in the public debate and since budgets are constrained by the weight of public debts.

3.5 Appendices

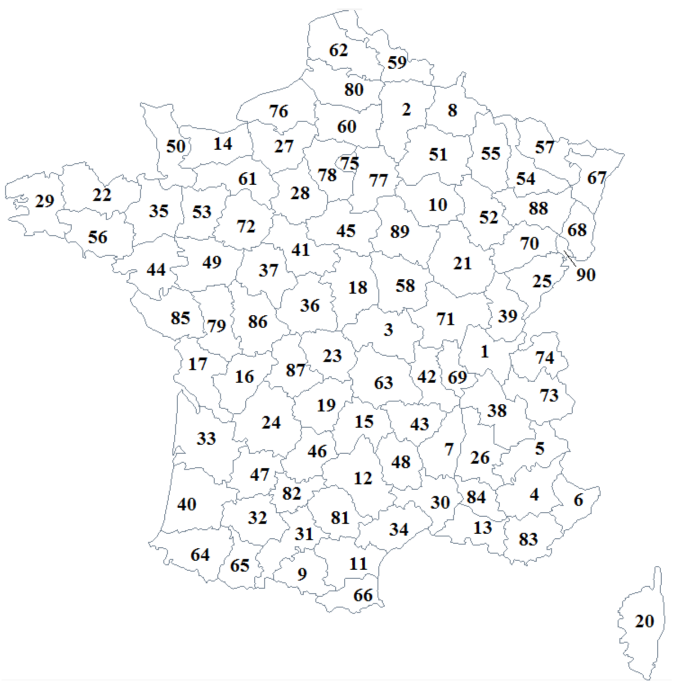
3.5.1 Publications Used to Compute Departmental Incomes

Year	Name of publication	Location of publication
1922–1928	RSRID 1930	
1929–1930	RSRID 1931–1932	
1931–1974	RSRID 1933 to RSRID 1975	
1975	ASDGI 1976	
1978–1985 (1)	ASDGI 1979 to ASDGI 1986	Archives of Finance Ministry, Savigny-le-Temple
1986–1989 (2)	ASDGI 1987, 1988, 1990, 1991	
1990–1998 (3)	ASDGI 1992 to ASDGI 2000	
1999–2000 (1)	ASDGI 2001 and ASDGI 2002	
2001–2002	Bonnet and Sotura (2018)	
2003–2014	ASDGI 2004 to ASDGI 2015	`https://www.impots.gouv.fr/portail/statistiques`

Notes: RSRID: *Renseignement Statistiques Relatifs aux Impôts Directs*; ASDGI: *Annuaire Statistique de la Direction Générale des Impôts.*

(1) No data for Imposable Income.
(2) No data for Imposable Income and Imposable Tax Units.
(3) Data for Imposable Income come from Bonnet and Sotura (2018).

3.5.2 Map of the 90 French *Départements* in 1967



1 Ain
2 Aisne
3 Allier
4 Alpes (Basses)
5 Alpes (Hautes)
6 Alpes Maritimes
7 Ardèche
8 Ardennes
9 Ariège
10 Aube
11 Aude
12 Aveyron
13 Bouches du Rhone
14 Calvados
15 Cantal
16 Charente
17 Charente Maritime
18 Cher
19 Corrèze
20 Corse
21 Côte d'Or
22 Côtes du Nord
23 Creuse
24 Dordogne
25 Doubs
26 Drôme
27 Eure
28 Eure et Loir
29 Finistère
30 Gard
31 Garonne (Haute)
32 Gers
33 Gironde
34 Hérault
35 Ille et Vilaine
36 Indre
37 Indre et Loire
38 Isère
39 Jura
40 Landes
41 Loir et Cher
42 Loire
43 Loire (Haute)
44 Loire Inférieure
45 Loiret
46 Lot
47 Lot et Garonne
48 Lozère
49 Maine et Loire
50 Manche
51 Marne
52 Marne (Haute)
53 Mayenne
54 Meurthe et Moselle
55 Meuse
56 Morbihan
57 Moselle
58 Nièvre
59 Nord
60 Oise
61 Orne
62 Pas de Calais
63 Puy de Dome
64 Pyrénées (Basses)
65 Pyrénées (Hautes)
66 Pyrénées Orientales
67 Rhin (Bas)
68 Rhin (Haut)
69 Rhône
70 Saône (Haute)
71 Saône et Loire
72 Sarthe
73 Savoie
74 Savoie (Haute)
75 Seine
76 Seine Inférieure
77 Seine et Marne
78 Seine et Oise
79 Sèvres (Deux)
80 Somme
81 Tarn
82 Tarn et Garonne
83 Var
84 Vaucluse
85 Vendée
86 Vienne
87 Vienne (Haute)
88 Vosges
89 Yonne
90 Territoire de Belfort

Notes : Numbers used in Bonnet (2018a)'s database. Corse is unified in this classification.

3.5.3 Methodology Used to Compute "Mortality Adjusted Incomes"

This methodology is based on d'albis and Bonnet (2018a)'s paper. To compute the "mortality adjusted incomes", I start from the program of a representative agent born in *département* i and in t. This agent maximizes:

$$\int_t^\infty e^{-\rho(z-t)} s_{t,i}(z-t) u(c_{t,i}(z)) dz, \tag{3.5.1}$$

under the budget constraint:

$$\int_t^\infty e^{-r(z-t)} s_{t,i}(z-t) y_{t,i}(z) dz = \int_t^\infty e^{-r(z-t)} s_{t,i}(z-t) c_{t,i}(z) dz. \tag{3.5.2}$$

Variables $c_{t,i}(z)$ and $y_{t,i}(z)$ represent consumption and income at date $z \geq t$, and $s_{t,i}(z-t)$ the probability of surviving to age $z-t$. Also, ρ and r are the subjective discount rate and interest rate respectively. Assuming $r = \rho$ and $y_{t,i}(z) = y_{t,i}$, consumption is constant and equal to income. The intertemporal utility, denoted $v(y_{t,i}, s_{t,i})$, can be written as:

$$v(y_{t,i}, s_{t,i}) = u(y_{t,i}) a(s_{t,i}), \tag{3.5.3}$$

which corresponds to the product of the utility of income, $u(y_{t,i})$, and the value of an annuity calculated using survival function $s_{t,i}$,

$$a(s_{t,i}) = \int_t^\infty e^{-r(z-t)} s_{t,i}(z-t) dz. \tag{3.5.4}$$

To compute the "mortality adjusted income", Fleurbaey and Gaulier (2009) calculate a willingness to pay, denoted $x(s_{t,i}, s_{t,i^*})$, by comparing for a given date the life expectancy in country i, with that of the country with the highest life expectancy, denoted i^*. In the case of French *départements,* this willingness to pay corresponds to the reduction in income an individual in *département* i would be willing to accept to enjoy the life expectancy in *département* i^*.

$$v(y_{t,i}, s_{t,i}) = v(y_{t,i} - x(s_{t,i}, s_{t,i^*}), s_{t,i^*}), \tag{3.5.5}$$

where $y_{t,i} - x(s_{t,i}, s_{t,i^*})$ corresponds to the "mortality adjusted income". By using Equation (3.5.3) to simplify Equation (3.5.5), one have:

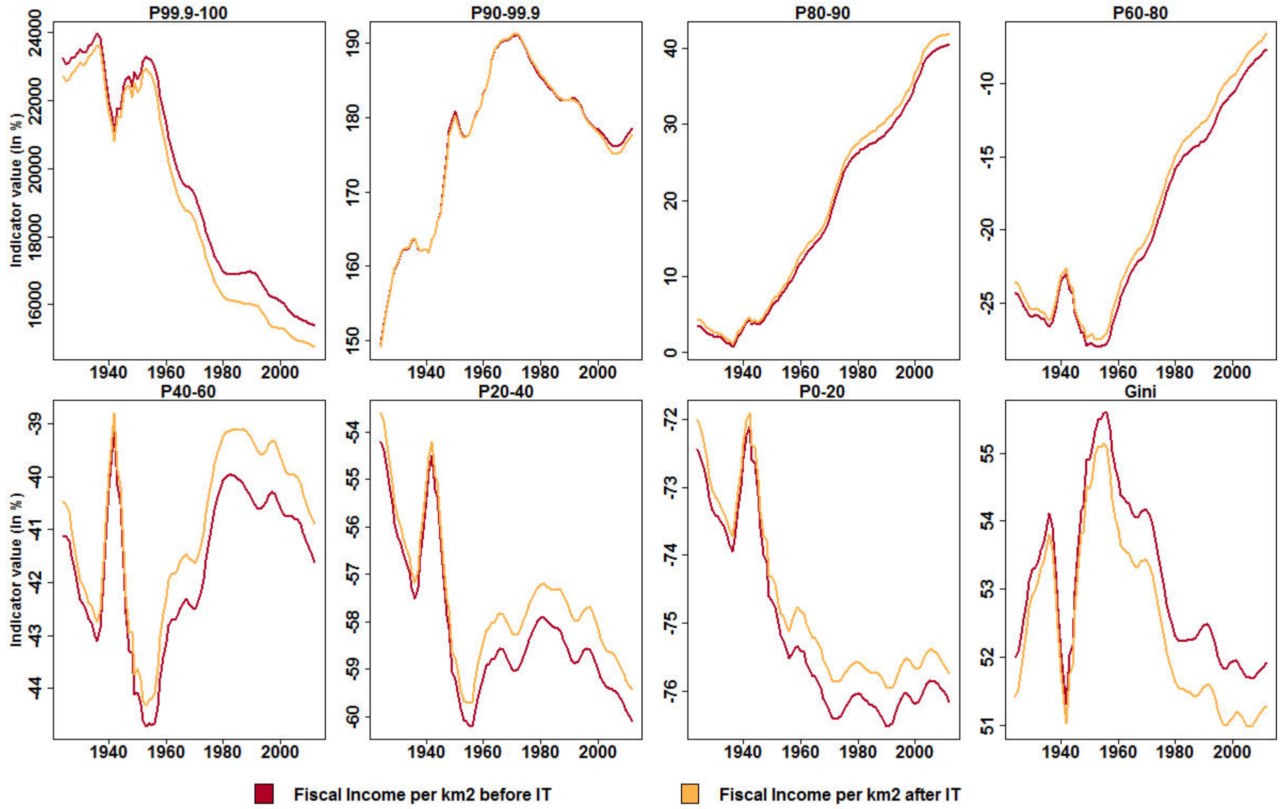
$$y_{t,i} - x(s_{t,i}, s_{t,i^*}) = u^{-1}\left(\frac{u(y_{t,i})\, a(s_{t,i})}{a(s_{t,i^*})}\right). \tag{3.5.6}$$

Like Becker et al. (2005), I use a CRRA function:

$$u(c) = \frac{c^{1-\frac{1}{\gamma}}}{1-\frac{1}{\gamma}} + \alpha, \tag{3.5.7}$$

and the same parameters: $\gamma = 1.25$, $\alpha = -16.2$ et $r = 0.03$. Finally, I obtain the "mortality adjusted incomes" for each *département* and each year.

3.5.4 Spatial Inequalities of Income Densities



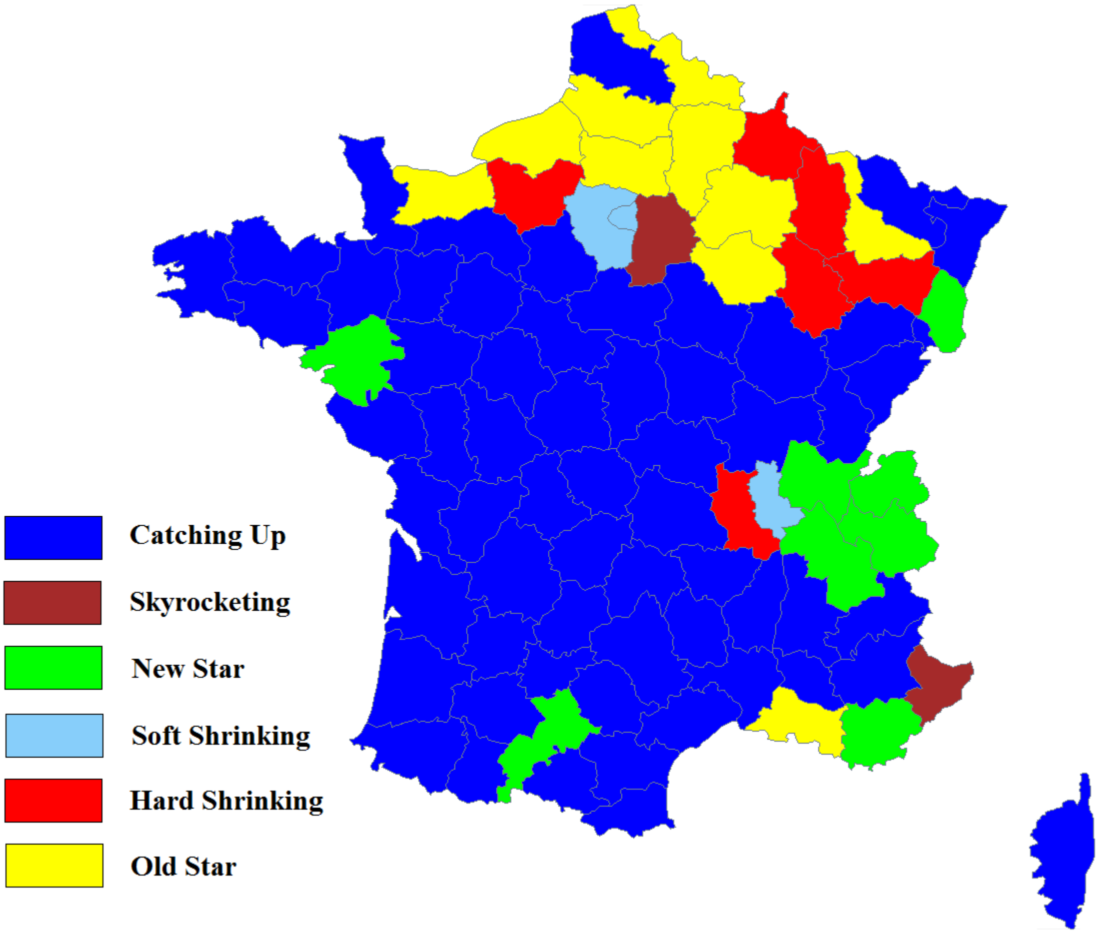
Notes: Income density in a *département* is the quotient between total income and number of km^2. P99.9-100 means the share of fiscal income earned by the 0.1% of km^2 with the highest density of incomes (compared with a spatial uniform density of income*)*. "IT" means Income Tax. All inequality indicators are weighted by the number of km^2. Inequality indicators are smoothed over a 5-year period. Sample includes 90 *départements*.

3.5.5 Shrinking Regions: A Classification

Département	Classification	Mortality Adjusted Income per adult (in % of the national mean)		
		1922	2014	Variation
Ain	New Star	76	108.1	32.1
Aisne	Old Star	120.5	82.1	-38.4
Allier	Hard Shrinking	86.1	84.3	-1.8
Alpes (Basses)	Catching up	59.9	89.6	29.7
Alpes (Hautes)	Catching up	55.8	93.4	37.6
Alpes Maritimes	Skyrocketing	101.6	112.2	10.6
Ardèche	Catching up	64.4	88.3	23.9
Ardennes	Old Star	106	80.7	-25.3
Ariège	Catching up	66.3	78.7	12.4
Aube	Old Star	108.1	89.5	-18.6
Aude	Catching up	71.3	79.6	8.3
Aveyron	Catching up	56.7	85.8	29.1
Bouches du Rhone	Catching up	89.4	98.6	9.2
Calvados	Hard Shrinking	95.5	94	-1.5
Cantal	Catching up	63.1	82.1	19
Charente	Catching up	84.3	87.3	3
Charente Maritime	Catching up	84.1	93.1	9
Cher	Catching up	86.2	87.3	1.1
Corrèze	Catching up	63.1	87.1	24
Corse	Catching up	31.1	82.5	51.4
Côte d'Or	Old Star	103.4	97.6	-5.8
Côtes du Nord	Catching up	57.2	90.3	33.1
Creuse	Catching up	63.2	75.5	12.3
Dordogne	Catching up	69.8	82.2	12.4
Doubs	Catching up	88	98.8	10.8
Drôme	Catching up	80.7	93.4	12.7
Eure	Hard Shrinking	99.2	93.3	-5.9
Eure et Loir	Hard Shrinking	98.3	98	-0.3
Finistère	Catching up	61.7	93	31.3
Gard	Catching up	71.1	86.3	15.2
Garonne (Haute)	New Star	83.8	102.9	19.1
Gers	Catching up	63.1	85.2	22.1
Gironde	Old Star	101.4	99.2	-2.2
Hérault	Catching up	81	91.4	10.4
Ille et Vilaine	Catching up	72	96.4	24.4
Indre	Hard Shrinking	84.2	83.3	-0.9
Indre et Loire	Catching up	92.9	96.3	3.4
Isère	New Star	93	102.9	9.9
Jura	Catching up	80.1	93.9	13.8
Landes	Catching up	80	91.8	11.8
Loir et Cher	Catching up	91.4	92.9	1.5
Loire	Hard Shrinking	90.8	89	-1.8
Loire (Haute)	Catching up	60.5	85.4	24.9

Département	Classification	Mortality Adjusted Income per adult (in % of the national mean)		
		1922	2014	Variation
Loire Inférieure	New Star	92.5	100.8	8.3
Loiret	Old Star	103.9	99.2	-4.7
Lot	Catching up	61.6	87	25.4
Lot et Garonne	Catching up	76.3	82.9	6.6
Lozère	Catching up	52.3	80.4	28.1
Maine et Loire	Catching up	90.4	91.2	0.8
Manche	Catching up	70.3	87.8	17.5
Marne	Old Star	117.7	97.4	-20.3
Marne (Haute)	Hard Shrinking	92	82.1	-9.9
Mayenne	Catching up	69.6	89.2	19.6
Meurthe et Moselle	Old Star	101.4	90	-11.4
Meuse	Hard Shrinking	95.6	82	-13.6
Morbihan	Catching up	59.8	94.3	34.5
Moselle	Catching up	72.1	89.9	17.8
Nièvre	Hard Shrinking	85.3	80.4	-4.9
Nord	Old Star	113.3	86.4	-26.9
Oise	Old Star	113.3	98	-15.3
Orne	Catching up	80.7	82	1.3
Pas de Calais	Catching up	77	80	3
Puy de Dome	Catching up	84.4	94	9.6
Pyrénées (Basses)	Catching up	84.8	97.6	12.8
Pyrénées (Hautes)	Catching up	70.3	86.8	16.5
Pyrénées Orientales	Catching up	73.7	83.6	9.9
Rhin (Bas)	Catching up	93	99.1	6.1
Rhin (Haut)	New Star	93.6	105	11.4
Rhône	Soft Shrinking	130.8	107.8	-23
Saône (Haute)	Catching up	80	84.2	4.2
Saône et Loire	Catching up	80.5	89.1	8.6
Sarthe	Catching up	88	90.2	2.2
Savoie	New Star	81.9	105.3	23.4
Savoie (Haute)	New Star	77.1	130	52.9
Seine	Soft Shrinking	180.9	129.3	-51.6
Seine Inférieure	Old Star	104.2	91.5	-12.7
Seine et Marne	Soft Shrinking	113	107.8	-5.2
Seine et Oise	Soft Shrinking	132.8	130.7	-2.1
Sèvres (Deux)	Catching up	80.2	87.3	7.1
Somme	Old Star	104.2	85.2	-19
Tarn	Catching up	80.1	86.9	6.8
Tarn et Garonne	Catching up	71	84	13
Var	New Star	79.9	101.2	21.3
Vaucluse	Catching up	76.9	91.1	14.2
Vendée	Catching up	78.5	91.7	13.2
Vienne	Catching up	83.1	88	4.9
Vienne (Haute)	Catching up	77.8	86.5	8.7
Vosges	Hard Shrinking	91.5	84.3	-7.2
Yonne	Hard Shrinking	90.9	89.9	-1
Territoire de Belfort	Hard Shrinking	93	90.2	-2.8

3.5.6 Departmental Classification According to Relative Income per Adult



Notes: Departmental classification according to the variation of relative income per adult between 1922 and 2014. Sample includes 90 *départements*.

Chapter 4

Beyond the Exodus of May-June 1940: Internal Flows of Refugees in France

Abstract

The Exodus of May-June 1940 remains in the French memory a traumatic episode. However, the internal population flows due to the war did not stop in 1940. This paper analyzes these internal flows of refugees in France during the whole Second World War. To do so, I present a method for estimating intercensal local populations from three kinds of statistical data, namely intercensal national populations, local census populations, and intercensal local births and deaths. I use this method to estimate female populations in the French *départements* between 1939 and 1946, and finally calculate the yearly departmental migratory flows. I show that the Exodus of May-June 1940 was only a first step in the massive flows observed in France at that time. Between 1940 and 1941 a large part of the population was displaced from the North to the South of the country. In detail, *Ardennes* was the most affected by departures: in 1941, 60% of the population counted in 1939 were missing. I also show that the *"ligne de démarcation"* represented, until 1943, a spatial boundary between the *départements* which welcomed refugees and those in which the population left. This result proves that this fictitious border did not prevent people from moving. Finally, I show that France kept the scars of war in 1946: the North-West and the North-East did not see all of their refugees come back. At the opposite, the population of the South-West, a land of refugees during the war, increased between 1939 and 1946 because of these internal flows of refugees.

4.1 Introduction

In April 1940, France was at war with Germany. Yet, little has changed in the life of the French during the "Phoney war". Children continued to attend classes, while non-mobilized men and women worked in factories or fields. However, two months later, the German army paraded in Paris' streets. The French army, which resisted for more than four years between 1914 and 1918, was defeated in a few weeks. Events followed quickly. First there was the invasion of Belgium and the Netherlands during the "18-day campaign", which forced hundreds of thousands of Belgians and Dutch to flee to the South. Then the *Sedan*'s breakthrough occured, while the General Staff of the Armies thought *Ardennes'* crossing was impossible. There was finally the defeat during the Battle of France: the French and English soldiers, enclosed in the Dunkirk pocket, only owed their salvation to the mobilization of the United Kingdom's fleet ("Dunkirk evacuation"). The advance of the German troops, inexorable during June 1940, led a large part of the population to the exodus from the North to the South of France. It has affected both the youngest and the oldest, women and non-mobilized men.

In this paper, I analyze internal flows of refugees on an annual basis during the Second World War. Indeed, if the Exodus of May-June 1940 remains in the collective mind a traumatic event, the population flows inside the country were also important during the rest of the war. They have not received the same attention yet: one need both accurate data and a specific methodological protocol to estimate these population flows. As a result, there are two major contributions in this paper.

The first is methodological. I propose a method based on three kind of data to estimate historical local populations for each year of an intercensal period: intercensal national populations, local census populations, and intercensal local births and deaths by cause. This method is based on the distinction between "asymmetrical mortality" – affecting only specific parts of the national territory - and "shared mortality". By assuming that the relative rate of shared mortality (ratio between departmental and national rates) has evolved linearly during the intercensal period, I can distribute the national population among all the *départements*. As far as I know, it's the first time that a method estimates historical local populations on an annual basis[1]: over the recent period, the study of internal migrations is mostly done by intercensal period. The method is based on a question asked to each individual who was surveyed about their place of residence in the previous census. Examples include Courgeau (1978), Baccaïni et al (1993), Baccaïni (1999) and Courgeau and Lelièvre (2004).[2] Nevertheless, none of these works estimated internal population flows on an annual basis.

This statement is the same according to the earlier period, but other statistical sources are mobilized. Kesztenbaum (2014) used military registers and explained how the residences of soldiers are

[1] The methodological protocol used by the Human Mortality Database (Wilmoth et al., 2007) to calculate lifetables over a long period is an example. These lifetables requires the populations on January 1^{st} for each year. For each intercensal period, apparent migration flows are calculated in two stages. First, by computing the net flow between the two censuses: to do so, one has to compare the census population in the second census with the census population in the first census from which are substracted the death of the cohort between the two censuses. Then one has to assume that the probability of migration is proportional to the elapsed time. This assumption is irrelevant during the Second World War. Indeed, the probability of migration was much greater between 1939 and 1940 than between 1937 and 1938 in the invaded *départements*.

[2] Using details in census data, population flows can also be disaggregated by sex or work category, as did Dean (1988). Other authors such as Pumain (1986) highlighted preferential flows of migration between French regions.

known between age 20 to 46, which allows to track their movements. Houdaille (1971) already used these registers to estimate inter-municipal and inter-departmental migrations. Family surveys are also an important source for tracking these migrations. The TRA survey (Bourdieu et al., 2014) followed generations of families whose last name begins with these three letters. This survey gave birth to many works: among them, Rosental (1999) studied women's migrations in the 20th century. Finally, collaborative genealogy data is an increasingly used source for tracking internal migrations. Brunet and Bideau (2000) already explained the advantages of these data twenty years ago. They are compiled by thousands people who wish to know their family history. They make it possible to follow precisely the path of each one. Among other things, they give interesting data by territory, according to mortality and birth rates. Charpentier and Gaillic (2018) used them for example to estimate migration flows to Paris in the 20th century.

The second contribution of this paper is analytical. I use this new method to compute the annual populations of each *département* between 1939 and 1946, and track the internal flows of refugees over this period. Until now, works of historians according to this issue have been based on paper archives, photos, and administrative documents. They have focused mainly on the massive flows of the May-June 1940 Exodus. Because of its intensity in a very short period, it still marks the French collective memory today. Jackson (2003), Diamond (2008) and Alary (2013) retraced this flash exodus. They showed how the French government allocated a reception *département* in the South and the West to each Northern and Eastern *départements*. Thus, before the invasion of German troops, the population of *Alsace-Moselle* was evacuated and sent to these regions. They also reported how the flow of people fleeing to the South disrupted the arrival of French reinforcements during the *Ardennes*' breakthrough. They finally showed the scenes of generosity in the host *départements*, but also annoyance in front of these one-day-migrants that changed habits and put a risk on the available food reserves. The gradual return of refugees is also described: it was difficult to achieve because of the disorganization and destruction of railways. Jackson (2004) and Alary (2006) shed light to the everyday life of the French during the German occupation; they emphasized the difficult cohabitation with the enemy soldiers, especially the Italians in the South-East of the country. Alary (2006) also detailed the economic disorganization due to the demarcation line (which separated until 1943 the German-dominated North from the South): the production was disrupted as raw materials were on one hand and processing plants on the other.

Works of historians based on archival documents are important to trace the episode script; quantitative history is however also useful. For example, according to the recent period, Schmeidl (1997) analyzed the reasons for mass migration during conflicts. This is also the case in Salehyan and Gleditsch (2006)'s work: they showed how the population flows caused by wars can destabilize the host territories. One would think that this kind of work cannot be replicated for a conflict as old as the Second World War, because of the destruction of the raw statistics. Yet, public statistics functioned during the war: births were recorded in 86 of the 90 metropolitan *départements*, as were deaths by age group and cause of death. These data did not disappeared, and are freely available in official publications.

By computing these internal flows of refugees, I show that they were dramatic in the North-

East, especially at the expense of *Ardennes* and *Meuse*. In *Ardennes*, more than 40% of the female population fled between 1939 and 1940, and more than 20% between 1940 and 1941. The return of these refugees to the North-East was particularly slow, mainly because of the creation of the North-East forbidden zone. Until December 1941, the refugees could not return within this geographical area, so that the German populations could gradually colonize it. Moreover, I show how the refugee places changed during the war: they were first in the Center and the West in 1940, then in the South beyond the North-South demarcation line until 1943. At the end, refugees concentrated mainly in the rural central *départements*. In particular, these results show that the North-South demarcation line marked until 1943 a boundary between the *départements* whose population increased due to migratory flows, and the *départements* in which the population decreased. While the Germans wanted this new border to be hermetic, it did not stop these flows of people.

The rest of this paper is organized as follows. I present data and methods used to calculate internal migrations during the Second World War in Section 2. In Section 3 I present the results for each year according to migratory flows. The fourth section concludes.

4.2 Data and Methods

4.2.1 Departmental data

Two statistical sources are used at the departmental level. The first is Bonnet (2018)'s database, which focuses on departmental mortality in France during the 20th century and is based on the exhaustive collection of both vital statistics and census data by sex. I extract two types of information from this study. The first is the population recorded in 1936 and 1946 censuses, which surround the Second World War. These statistics are available for each sex and *département*. The second are vital statistics, namely the number of births and deaths yearly recorded between 1936 and 1946. They are available for each sex and each *département* except *Moselle, Bas-Rhin* and *Haut-Rhin* between 1939 and 1945, and *Corse* in 1942 and 1943. Note that I use these statistics as they were collected in the archives. Therefore, the methods used by Bonnet (2018) to calculate the departmental lifetables have no impact on the results presented in this study.

The second statistical source used is an exclusive one. In addition to the deaths recorded each year, offical publications provide information about the cause of death for the 1936 and 1940–1946 years. Unfortunately, statistics for the 1937-1939 period were destroyed during the war. These deaths by cause are available for each *département*. They are also available by sex for the 1936 and 1942-1946 years. The International Classification of Diseases used at that time to distinguish causes of death remained the same throughout the period. The only exception is deaths due to car accidents, isolated in the 1944 classification. Appendix 4.5.2 produces this nomenclature.

Note that the departmental totals of deaths, whether classified by age in Bonnet (2018)'s database or by cause of death, coincide.

4.2.2 Framework of the Study

This study concerns the female population for two reasons. The first is military deaths, which are not included in the raw materials of the *Statistique Générale de la France*. In addition, the indirect sources used by Bonnet (2018) do not provide the home *département* of the deaths but the *département* of birth. Moreover, women's migrations are in principle less constrained than those of men. The latter had first to join the battlefields of the North and East, before being made prisoners and sent to camps in Germany. For those remaining in France, a large number were also sent to Germany because of the *"Service du Travail Obligatoire"*. With this restriction, the reliability of this study should not be questioned.

The spatial framework does not cover the metropolitan territory as it is defined today. I excluded *Moselle, Bas-Rhin, Haut-Rhin* (these three *départements* are *Alsace-Lorraine*) and *Corse* because their data are unavailable between 1939 and 1945. Finally, my sample includes 86 *départements* (see map in Appendix 4.5.1). To estimate the total population of the 86 *départements* for each year between 1936 and 1946, I use Vallin and Meslé (2001)'s data. In their study of the French age-specific mortality during the 20th century, Vallin and Meslé (2001) estimated the national population on January 1^{st} of each year using census and vital statistics data. However, their definition of national territory changes along the period: data do not include *Moselle, Bas-Rhin, Haut-Rhin* and *Corse* when the corresponding statistics are missing. Thereby, I have to ensure the consistency of data between departmental and national sources. Table 4.1 (Column 2) presents the national populations in Vallin and Meslé (2001)'s database. From 1936 to 1939 and in 1946, these data cover the metropolitan territory as we know it today. They do not include *Alsace-Lorraine* from 1939 to 1942, as well as *Corse* from 1943 to 1945. For the key years, data are available according to the old and the new definition of the national territory, which makes it possible to calculate the weight of the missing *départements* (Table 4.1, in italics). For example, before 1938, *Alsace-Lorraine* accounted for 4.60% of the total. Using this information, I estimate the population at January 1^{st} for the 86 *départements* of my study (Column 5). Subsequently, the term "national" will refer to this territory's definition. Note that average populations of each year are calculated by averaging populations of two successive years.

4.2.3 Estimation of Annual Departmental Populations and Internal Migrations

Methodological framework

To estimate annual internal migration flows during an intercensal period, one must know the annual populations of each *département*. Assuming that I know the yearly national populations, the goal is to allocate this population between each of the *départements*. To do so, I start from the crude mortality rate $q_{i,t}$:

$$q_{i,t} = \frac{D_{i,t}}{P_{i,t}}, \tag{4.2.1}$$

Table 4.1: Estimates of national population of women at january 1^{st}, 1936–1946

Year	Pop. January 1^{st} Vallin and Meslé (2001)	% Corse	% Alsace-Lorraine	Pop. January 1^{st} 86 *départements*
1936	21,390,450[1]	0.60%	4.60%	20,277,826
1937	21,406,000[1]	0.60%	4.60%	20,292,568
1938	21,422,700[1]	0.60%	*4.60%*	20,308,399
1939	20,439,918[2]	0.60%		20,317,564
1940	20,520,230[2]	0.60%		20,397,395
1941	20,235,330[2]	0.60%		20,114,200
1942	20,165,130[2]	*0,60%*		20,044,420
1943	20,009,630[3]			20,009,630
1944	19,981,430[3]			19,981,430
1945	19,936,730[3]			19,936,730
1946	21,006,118[3]	*0.60%*	*4.39%*	19,957,902
1947	21 147,621[1]	0.60%	4.39%	20,092,344

Notes: Population of women at January 1^{st} calculated by Vallin and Meslé (2001) according to different definitions of the national territory. Italic numbers are the *départements* weights in the national population for the last known years. "86 *départements*" specification excludes *Moselle*, *Bas-Rhin*, *Haut-Rhin* and *Corse.*
1: Population of women with 90 *départements*; 2: Population of women without *Moselle, Bas-Rhin* and *Haut-Rhin*; 3: Population of women without *Moselle, Bas-Rhin, Haut-Rhin* and *Corse*

with D the total of deaths, P the average population, t the year and i the *département*. In the same way, one can write this relationship at the national level:

$$q_{nat,t} = \frac{D_{nat,t}}{P_{nat,t}}. \tag{4.2.2}$$

I assume that, at dates of the first and the second censuses (T_1 and T_2) of the intercensal period, the average population of the year is equal to the recorded population. Thereby both departmental and national mortality rates are available on these dates and I deduce the relative mortality rates θ:

$$\theta_{i,t} = \frac{q_{i,t}}{q_{nat,t}} = \frac{\frac{D_{i,t}}{P_{i,t}}}{\frac{D_{nat,t}}{P_{nat,t}}} \quad t = \{T_1, T_2\}. \tag{4.2.3}$$

Nevertheless, I can not make a linear approximation of these values between T_1 and T_2 to deduce the departmental mortality rates and consequently the populations. I would assume that local mortality rates have evolved in the same way as death rates at the national level. Therefore, two types of mortality must be distinguished. The first, which I call "shared mortality", is a mortality whose national and local evolutions are synchronized. The second, which I call "asymmetrical mortality", is a mortality whose national and local evolutions are out of synchronization. From this distinction, I rewrite Equation (4.2.3) as follows in Equation (4.2.4). Let call q^* the mortality rates that do not account for asymmetric mortality. Thus, for T_1 and T_2:

$$\theta^*_{i,t} = \frac{q^*_{i,t}}{q^*_{nat,t}} \quad t = \{T_1, T_2\}. \tag{4.2.4}$$

I assume that θ^* evolved linearly between T_1 and T_2. In this way, I can estimate the relative mortality rates for each *département* and intercensal year, called $\hat{\theta^*_{i,t}}$. Using $\hat{\theta^*_{i,t}}$, I compute the average populations for each *département* and each year of the intercensal period ($\hat{P_{i,t}}$):

$$\hat{P_{i,t}} = \frac{D^*_{i,t}}{q^*_{nat,t} \times \hat{\theta^*_{i,t}}} \quad t = \{T_1, T_1 + 1, ..., T_2\} \, . \tag{4.2.5}$$

Using yearly births B and deaths D, I get apparent migratory flows M:

$$\hat{M_{i,t}} = (\hat{P_{i,t}} - \hat{P_{i,t-1}}) - (D_{i,t} - B_{i,t}). \tag{4.2.6}$$

To conclude, it should be noted that these estimated migratory flows are different from the intercensal migratory flows. Indeed, a census is a picture at time t of the population living in one *département*. The population I estimate in this study is not that of a census: it is the average population living in the *département* during the past year. To understand it, let us take the example of a territory on which 1000 individuals live continuously during the year *t*. The intrinsic growth is zero. The whole population leaves the territory during 10% of the year *t + 1* and then comes back for the rest of the year. Consequently, the estimated population in t + 1 is 900 individuals, i.e. a migratory flow of 100. Thus, this method does not allow precise isolation of flash migration flows such as that of the May-June 1940 Exodus, but reveals them through two factors: their intensity and their duration.

Application to the Second World War

For this study, I calculate populations and internal migrations during the 1936–1946 intercensal period. The central element of the method I have proposed is to disentangle the "asymmetrical mortality from the "shared mortality" in order to keep only the latter. To do so, I use statistics by cause of death. An examination of the nomenclature used to classify the deaths allows to isolate the 42^{th} category (43^{th} after 1944) "Violent or accidental death (except suicide and homicide)" in which are the civilian deaths due to wars. This category includes deaths from bombing, whether from German or Allied forces.

Table 4.2 presents some global statistics concerning the share of deaths from accidental or violent deaths in the total, for the . The weight of these deaths increased sharply during the war (Line 1), from 1.6% in 1936 to 3.5% in 1940 and 9.2% in 1944. This increase was not spatially the same: while the maximum was 2.9% in 2006, it reached a peak in 1944 with 41.6% in *Calvados* (Line 3). On the other hand, some regions remain isolated from the war: the minimum is constant, around 1% of total deaths (Line 2). This is the case of the South-West rural *départements* such as *Ariège* and *Lozère*. The maximum in territorial differentiation is reached in 1944, with a standard deviation of 6.5 percentage points (Line 4). Some historical events are visible in these statistics: in 1943, the share of violent deaths is reached in *Loire-Inférieure* (12.2%). This is explained by the Allied bombing of September, which killed 1,450 people in *Nantes*.

Figure 4.2.1 represents these shares of violent or accidental deaths for the 86 *départements* in 1940

Table 4.2: Share of violent or accidental deaths, 1936–1946

	1936	1937	1938	1939	1940	1941	1942	1943	1944	1945	1946
National	1.6%				3.5%	2.3%	2.4%	3.5%	9.2%	2.6%	2.1%
Minimum	0.7%				0.8%	0.8%	0.9%	1.2%	0.7%	0.7%	0.6%
Maximum	2.9%				9.5%	4.5%	4.0%	12.2%	41.6%	8.0%	3.5%
Standard deviation	0.4				2.0	0.6	0.6	1.8	6.5	1.0	0.6

Notes: Computations based on the population of women. “National” refers to the 86 *départements* of the sample (excluding *Corse, Moselle, Bas-Rhin* and *Haut-Rhin*). Violent or accidental deaths belong to the 42^{th} (43^{th} after 1944) category of the International Classification of Diseases in force at the time (See Appendix 4.5.2 for more details according to the nomenclature).

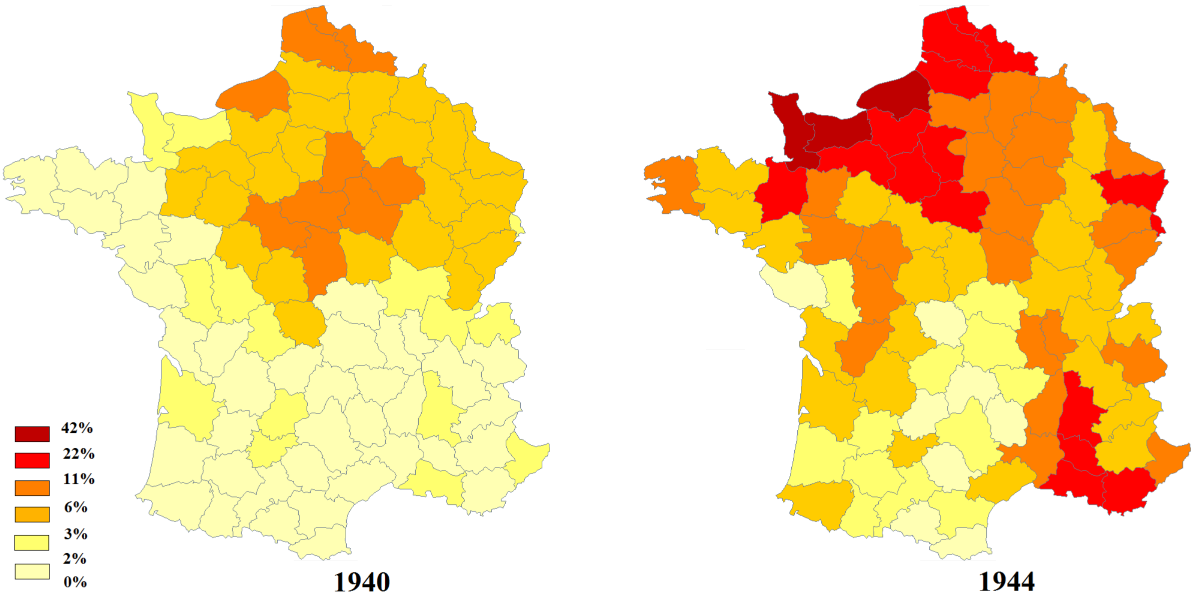
and in 1944, when this kind of mortality was the most important.[3] Table 4.2 reveals why I consider this mortality as asymmetrical. 1940 shows an excess mortality due to spatially concentrated warfare events: the whole North-West of the country was afflicted and violent death were maximum in the North (*Nord, Pas-de-Calais, Seine-Inférieure*) and the south-east of the Paris region (between 6 and 11%). The first area is explained by the fighting during the Battle of France, between May 10 and June 22, 1940. The second may be due to the Exodus of May-June 1940, which put on the roads population heavily exposed to raids of German aviation. The South and West were not affected: levels are similar to those of the pre-war period, between 0 and 2%. Indeed, German troops did not invade this area before the armistice. Values are much higher in 1944, with peaks reached in the northwestern and southeastern parts of the country. *Normandie*’s shores were particularly afflicted: share rose up to 40% in *Calvados* while violent deaths represented about 25% of the total in *Seine-Inférieure* and *Manche*. In contrast, *Massif Central*, *Vendée* and the Pyrenean border kept values close to those observed in 1936.

I subtract violent or accidental deaths from total deaths to keep only the “shared mortality”. The missing death-by-cause statistics for the 1937, 1938 and 1939 years may seem problematic. This is not the case for two reasons. The first is that this study is centered on the 1939–1946 years: these missing data are therefore problematic only in 1939. The second is that it is very likely that the departmental weights of these accidental or violent deaths are similar in 1939 to those of 1936: very low (around 1%) and homogeneous throughout the national area. Indeed, the Second World War really began in Western Europe on May 10, 1940 with the invasion of Belgium and Luxembourg by German troops. The *Sedan*’s breakthrough came a few days later. In Equation (4.2.4), this means that the numerator and the denominator are multiplied by the same value, that is the percentage of non-violent and non-accidental deaths. This leaves θ^* unchanged. In 1940 and 1941, the distribution by sex is not available concerning causes of death. However, this distribution is quite stable (around 30% for women, for each year and each *département*); I used 1942 values to tackle this issue.

Finally, for each year between 1939 and 1946 and for each of the 86 *départements*, I get three variables concerning the female population: the total of deaths and births as well as the average population.

[3]All the maps of this study have been realized with the Philcarto software `http://philcarto.free.fr/`

Figure 4.2.1: Share of violent or accidental deaths, 1940 and 1944



Notes: Computations based on the population of women. Violent or accidental deaths belong to the 42^{th} (43^{th} after 1944) category of the International Classification of Diseases in force at the time. Sample of 86 *départements* (excluding *Corse, Moselle, Bas-Rhin* and *Haut-Rhin*).

4.3 Results

4.3.1 The Global Consequences of the War

I first analyze the departmental variations of the female population between 1939 and 1946. I disentangle the apparent migratory movement from the intrinsic growth by using totals of births and deaths collected by Bonnet (2018). Table 4.3 presents the national population variations between 1939 and 1946 (in %) as well as a summary of the distribution of the departmental variations. Appendix 4.5.3 presents the estimated values for each of the 86 *départements* of the sample. Table 4.3 shows that the population decreased by 1.6% over the period at the national level (Column 1). This decrease was equally distributed between the intrinsic growth (0.7 point of contribution, Column 2), and the migratory movement (0.9 point, Column 3).

Table 4.3 also reveals that these national variations are far much lower than the departmental variations: the latter fall between +13.7% and -19.2% (a decrease 12 times greater than the one observed at national level). The distribution between the two components of these local variations shows that the intrinsic growth had only a small influence. At most, the population increased by 2.8% due to the surplus of births compared to deaths (*Pas-de-Calais*). Thus, migrations had the largest contribution: the rates of change ranged from -18.2% to +14.9%, with a standard deviation of 6.5 percentage points. The departmental populations evolved in a so differentiated manner because of internal population migrations.

Table 4.3: Variations of population by component, 1939–1946

		Total	Intrinsic growth	Migratory movement
Departmental distribution of population variations	Minimum	-19.2%	-4.6%	-18.2%
	Maximum	13.7%	2.8%	14.9%
	Median	-2.3%	-0.7%	-1.4%
	Standard deviation	6.4	1.5	6.5
National		-1.6%	-0.7%	-0.9%

Notes: Computations based on the population of women. "National" refers to the 86 *départements* of the sample (excluding *Corse, Moselle, Bas-Rhin* and *Haut-Rhin*). Population variations between the average population in 1939 and the recorded population in 1946. Intrinsic growth is the difference between births and deaths. Migratory movement is the difference between the whole variation and the intrinsic growth.

4.3.2 Annual Monitoring of Internal Population Migrations

Figures 4.3.1 to 4.3.4 map the variation of population due to migratory flows between one year and 1939 (in % of the July 1^{st}, 1939 population). They represent the cumulative of net migrations since the beginning of the war. Appendix 4.5.4 provides annual flows for each of the 86 *départements*.

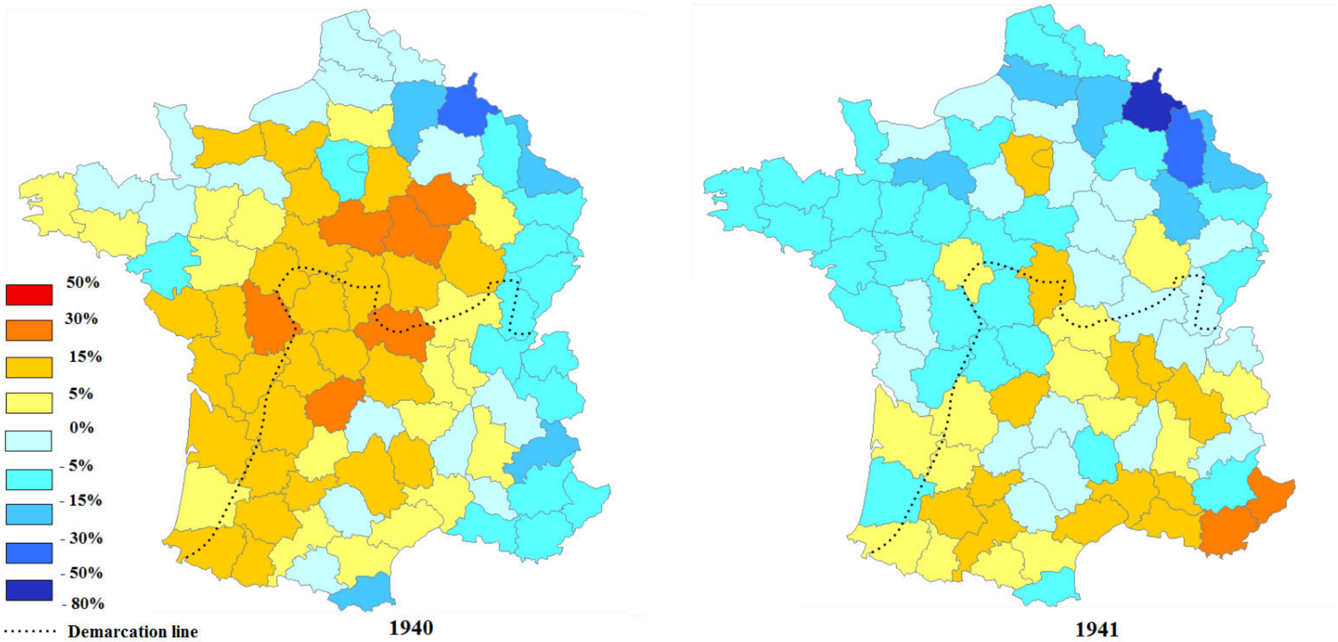
1940: the flight to the South-West

Figure 4.3.1 shows the well-known results according to the May-June 1940 Exodus. However, they cannot be directly compared to the figures of the displaced during these two months, since the method used does not estimate the infra-annual flows. This map shows how the population next to the eastern borders (Belgium, Luxembourg, Germany, Switzerland and Italy) as well as the ones living in *Seine* and *Seine-et-Oise* (corresponding to the Paris agglomeration) fled in the Center and the South-West of the country between 1939 and 1940. *Ardennes* lost 42% of its female population, *Meurthe-et-Moselle* 16%, *Hautes-Alpes* 17%. In a few days, when the capital was declared as "open city", a large part of the population left Paris and its suburbs. Nevertheless, the return flow was rapid: over the period, only 8% of the population is missing due to migrations. From *Marne* to *Pyrénées-Atlantiques* via *Corrèze*, all the South-West *départements* welcomed refugees.[4] The *départements* in the south of Paris welcomed the most of them. The population of *Loiret, Yonne* and *Aube* increased by about 16% due to these internal migrations. Also, one has to note the increase of 16% recorded in *Allier*, where the government settled after the Franco-German armistice of June 22, 1940. By aggregating these flows of population, it is the equivalent of 650,000 women who left more or less permanently the *départements* of North and East to flee in the Center and the South-West between 1939 and 1940. In addition, 60,000 women left to take refuge abroad.[5]

[4] The situation in *Pyrénées-Orientales* is different: the female population decreased by 17% between 1939 and 1940. This is explained by the flow of Spanish Republicans ("Retirada"), who fled Franco's taking of power. These refugees, concentrated on the Spanish border in 1939, transiently increased the population at that time. It then decreased because of their redistribution in the other *départements* of the South-West.

[5] Once again, the absolute values presented here cannot be directly compared to the figures of historians according to the May-June 1940 Exodus. Estimates relate to annual average flows, not sub-annual flows. For more details, see Section 4.2.3.

Figure 4.3.1: Population variations since 1939 due to migratory movement, 1940 and 1941



Notes: Computations made for the 86 *départements* of the sample (excluding *Corse, Moselle, Bas-Rhin* and *Haut-Rhin*), according to the population of women. Population variations between the average population in 1939 and the average population in 1940 and 1941. Migratory movement is the difference between the whole variation and the intrinsic growth. Intrinsic growth is the difference between births and deaths.

1941: no return for the northeastern refugees

The most striking statement of Figure 4.3.1 in 1941 is the dark blue spot in the North-East, and more particularly in *Ardennes* and *Meuse*. In 1941, 60% of the population was away from *Ardennes* because of migrations, 35% in *Meuse*. More broadly, 15 to 30% of the northeastern population left between 1939 and 1941. Thus, despite the Franco-German armistice, the flow of refugees did not seem to dry up. The explanation comes from the different restricted areas within the part of country controlled by the German. Beyond *Moselle, Bas-Rhin* and *Haut-Rhin* annexed by Germany, refugees cannot return to a wide area including *Doubs, Haute-Saône, Vosges, Meurthe-et-Moselle, Meuse, Ardennes* and the northern part of *Aisne* and *Somme*. Moreover, the German military command governed the population of *Pas-de-Calais* and *Nord*. In addition, one can see that the population of the West fled towards the free zone too: it is the case of *Bretagne, Normandie, Pays de la Loire* as well as a large part of the Center region, which hosted refugees from the May-June 1940 Exodus. 15% of their 1939 population left *Mayenne, Indre, Loir-et-Cher* or *Vendée*.

Two new refugee areas appeared in 1941: the Paris region and the South-East. For the first, the May-June 1940 Exodus was already far away. The population of *Seine* and *Seine-et-Oise* increased between 1940 and 1941 by 23% and 13%, respectively (in % of their population in 1939). *Île-de-France* thus was used as a refuge for the national population, even if difficulties in transportation probably made it a forced refugee area. Refugee flows were important in the South-East, especially near the Mediterranean coast. *Var* and *Alpes-Maritimes* welcomed the equivalent of 25% of their 1939 population between 1940 and 1941, that is 50,000 and 70,000 women, respectively.

The May-June 1940 Exodus is remembered as massive and risky because of military operations. Nevertheless, the 1940-1941 period concentrates the largest permanent population flows. Movements towards the capital and the South-East of the country put nearly 1,150,000 women on the roads, while another 100,000 went abroad. The whole represented about 7% of the total population.

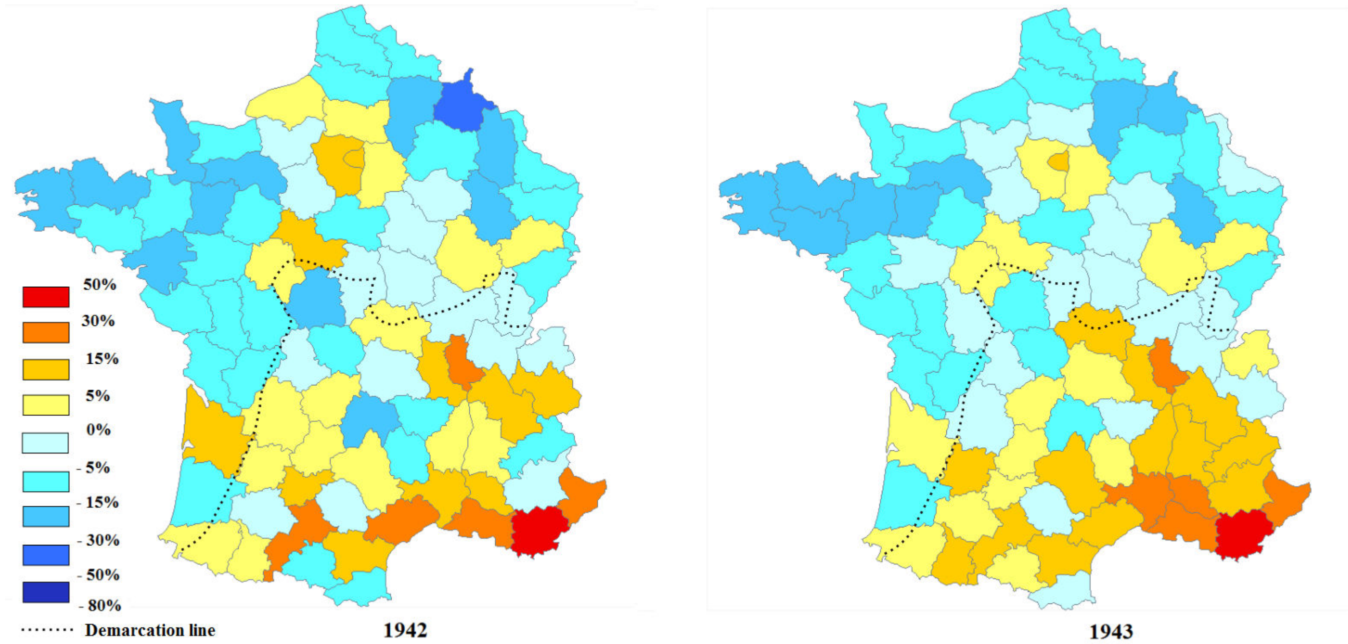
1942 and 1943: return of the northeastern refugees and departure towards the South

Figure 4.3.2 reveals population variations between 1939 and 1942 and between 1939 and 1943. Overall, migrations were three times smaller in 1941-1942 and 1942-1943 than those observed between 1940 and 1941. The net migration flow of women leaving their *départements* is around 450,000. The outward flow of the country dried up, since only 10,000 left the country. During this period, the populations living near the borders of the North-East came back, particularly in *Meuse* or *Ardennes*. The northeastern line that prohibited the return of refugees was gradually easing. Between 1941 and 1943, more than 60% of women who left these two *départements* came back to live there. The situation was similar for *Meurthe-et-Moselle*.

Concerning the rest of the country, Figure 4.3.2 reveals several interesting features. The first concerns the Paris region. The refugee flows reversed in *Seine* and *Seine-et-Oise*: 150,000 women left the region between 1941 and 1943. It is difficult to know precisely where they went. Nevertheless, it seems reasonable to assume that these women were both refugees from the North-East returning home and residents of the capital who also fled gradually beyond the demarcation line. The second concerns the North-West of the country, whose population continued to flee between 1941 and 1942: 8 to 10% of the 1939 population went to live elsewhere in *Calvados, Côtes-du-Nord, Finistère*, *Manche* or *Loire-Inférieure*. At this date, nearly 20% of the 1939 population is missing in these *départements* due to internal migrations. The Atlantic Wall which appeared in April 1941 may be one of the explanations. This area, 20 to 30 kilometers wide, from Dunkirk to *Pays Basque*, was prohibited to anyone without specific authorization.

The North-South gradient for population losses due to migrations was the most pronounced in 1943. The demarcation line, which disappeared since November 1942, left a strong imprint on the geography of migrations since the beginning of the war. From *Basses-Pyrénées* to *Jura* through *Charentes* and *Nièvre*, only a few *départements* north of this line present a gain in population due to migrations since 1939. This is the case of the Paris region, even if the flow was reversed since 1941. This is also the case of *Côte-d'Or* and *Haute-Saône*, which probably still hosted some refugees of the eastern frontiers who did not return to their homes. Concerning *Gironde*, although the stock of migrations since 1939 was still positive, the flow was strongly negative: between 1942 and 1943, 13% of the 1939 population emigrates, namely more than 60,000 women. For the remaining *départements*, a large South-East quarter welcomed refugees, with peaks reached in *Rhône* and along the Mediterranean coast. As such, 31% of its 1939 population settled in *Var*, 25% in *Alpes-Maritimes* and 22% in *Gard*.

Figure 4.3.2: POPULATION VARIATIONS SINCE 1939 DUE TO MIGRATORY MOVEMENT, 1942 AND 1943



Notes: Computations made for the 86 *départements* of the sample (excluding *Corse, Moselle, Bas-Rhin* and *Haut-Rhin*), according to the population of women. Population variations between the average population in 1939 and the average population in 1942 and 1943. Migratory movement is the difference between the whole variation and the intrinsic growth. Intrinsic growth is the difference between births and deaths.

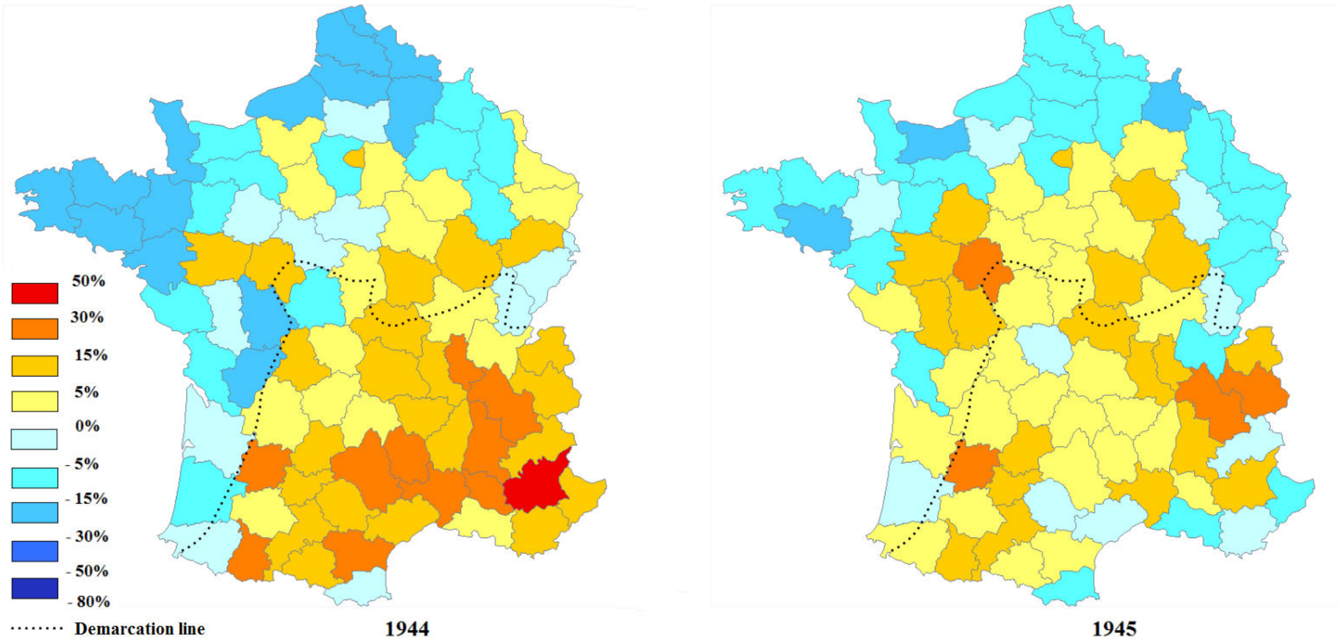
1944: the inland flight

The 1944 map (Figure 4.3.3) reveals a break from previous years, for several reasons. First, the demarcation line was not as marked as in 1943: the South-East still hosted a large number of refugees, but these refugees were progressively moving North. *Nièvre, Côte d'Or, Maine-et-Loire* and *Indre-et-Loire* turn yellow on the map. At the same time, the North of the country as well as *Bretagne* homogenize in blue, joined by all the *départements* in the south of Atlantic coast. Finally, refugees on the Mediterranean coast were starting to leave. Emigration from *Alpes-Maritimes, Var* and *Bouches-du-Rhône* represented more than 180,000 women, probably towards inland. As such, *Basses-Alpes* received between 1943 and 1944 the equivalent of 25% of its 1939 population. These values were about 15% for the nearby *départements* such as *Drôme, Lozère, Aveyron*, *Cantal* or *Haute-Loire*. These last two are particularly interesting insofar as they stayed away from the refugee wave although they were located beyond the demarcation line. These rural *départements* thus served as a refuge during this period. In aggregate terms, migration flows were picking up again: between 1943 and 1944, 575,000 women migrated to another department, and 25,000 to other countries.

These movements are probably explained by two reasons. The first is the occupation of the free zone by the Germans and Italians since November 1942 and by the Germans only since September 1943. The South-East was no longer a protected area for refugees. This is even more the case of the well urbanized Mediterranean coast, which saw the massive arrival of German troops. The second concerns the preparations for landings, whether on *Normandie*'s beaches (June 1944) or in *Provence* (August 1944). The populations surely fled the allied bombings that preceded them, in *Nantes*

(September 1943, 1,450 victims), but also in *Marseille* (May 1944, 1,750 victims). In their flight, the populations of the South-East coastal *départements* took refuge in the nearest rural *départements*, with no military targets.

Figure 4.3.3: Population variations since 1939 due to migratory movement, 1944 and 1945



Notes: Computations made for the 86 *départements* of the sample (excluding *Corse, Moselle, Bas-Rhin* and *Haut-Rhin*), according to the population of women. Population variations between the average population in 1939 and the average population in 1944 and 1945. Migratory movement is the difference between the whole variation and the intrinsic growth. Intrinsic growth is the difference between births and deaths.

1945: the gradual return of refugees settled in the South

Figure 4.3.3 shows much less diverse situations in 1945 than in 1944. Refugees gradually left areas which hosted the most of them during the Second World War to return to their home *départements*. Between 1944 and 1945, nearly 700,000 women participated in these reverse migrations, from South to North, making it the second highest total after 1940-1941.

In detail, a large part of the population of the Atlantic coast, *Bretagne* and the North of the country came back: *Loire-Inférieure* and *Ile-et-Vilaine* each welcomed 60,000 women between 1944 and 1945. This is almost all the missing due to migrations in *Ile-et-Vilaine*. This finding is similar in *Somme*, *Nord* and *Pas-de-Calais*, where incoming migratory flows were about 20,000 to 30,000 women. At the opposite, the South-East rural *départements* see their refugees went back to their home *départements*. Globally, it was the equivalent of 5 to 20% of their 1939 population which left *départements* close to the *Massif* Central. Values reached 9% in *Ardèche*, 13% in *Aveyron*, 18% in *Lozère*. Note that departures were also important along the Mediterranean coast. Beyond the return of refugees to their homes, the migratory movement between 1939 and 1945 had a negative contribution to the female population variation in *Var*, *Bouches-du-Rhône* and *Alpes-Maritimes*.

1946: the scars left by World War II

Figure 4.3.4 shows the final situation in 1946. This date is important for two reasons. Firstly because it is a census year: the population of 1946 is no longer estimated but recorded. Second, it is a year of gradual stabilization in France: even if the country was hurted by the fighting for the *Libération*, no more military operations were taking place on the metropolitan land: the Allied and German bombings stopped.

This map reveals a first global statement: the absence of bright colors (population difference between 1939 and 1946 due to the net migration more than ± 15%) shows that almost all the refugees returned, which was not true in 1944 and 1945 yet. In detail, the last major migratory flows left from the South-East and the Center-West. *Départements* such as *Isère, Basses-Alpes, Haute-Loire* or *Indre-et-Loire, Vienne* and *Haute-Vienne* saw the return of 10% of their 1939 population between 1945 and 1946. These migratory flows were mainly directed towards three partially depopulated areas: the North of the country – with values of 9% and 11% for *Nord* and *Pas-de-Calais* – *Normandie* which had been strongly affected by the landing – values close to 7% for *Calvados, Manche* and *Eure* – as well as the North-East – values of 11% for *Ardennes* or 7% for *Aisne*. The flows towards *Nord, Pas-de-Calais, Ardennes* and *Aisne* were about 200,000 women, that is 40% of the total flow over the period. Note that the net flow to foreign destinations was finally reversed, with the return of 40,000 women in France.

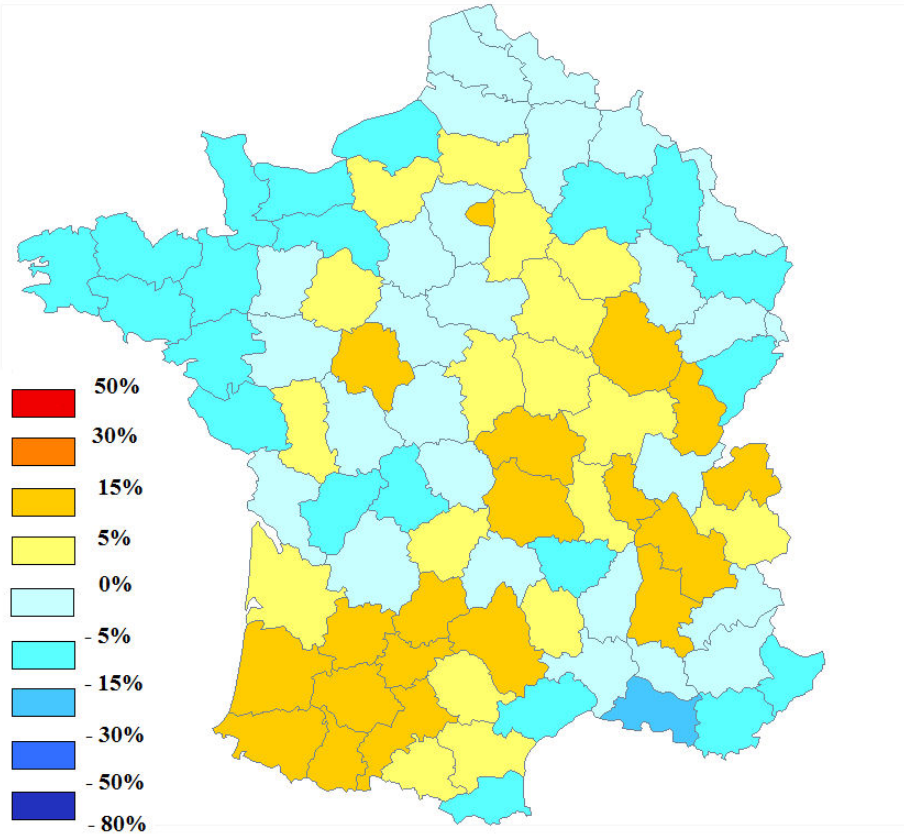
This map refines the analysis of the scars started in section 4.3.1. Even if a large part of the metropolitan territory recovered the population lost due to internal migrations, some areas reveal the war demographic stigma. This is the case of the North-East, which lost between 1939 and 1946 about 7% of its population due to internal migrations, with values above 10% in *Meuse* and *Vosges*. The situation is similar in *Normandie* and *Bretagne* where migrations caused the displacement of 8% of the population. The maximum values picked up in *Loire-Inferieure* (-14%), *Morbihan* (-13%) and *Seine-Inférieure* (-10%). Conversely, the South-West and some *départements* in the Central-East welcomed a population that never left. A particularly homogeneous area appears between *Aveyron*, *Landes* and *Hautes-Pyrénées*: over the period, the values are about 8%, with a maximum reached in *Lot-et-Garonne* (15%).

4.4 Conclusion

In this paper, I have first proposed a methodology to estimate local populations in an intercensal period. This methodology is based on the knowledge of the national population for each intercensal year, local populations for each census year and local vital statistics for each intercensal year. By using these vital statistics, I can distinguish the "shared mortality", which evolves in a synchronized manner at the local and national levels, from "asymmetrical mortality". To do so, the availability of deaths by cause is crucial.

With this methodology I have estimated the female populations of each *département* (outside *Corse, Moselle, Bas-Rhin* and *Haut-Rhin* for which the data are missing) and each year between 1939 and 1946. I have considered that the violent or accidental deaths had to be subtracted from

Figure 4.3.4: Population variations since 1939 due to migratory movement, 1946



Notes: Computations made for the 86 *départements* of the sample (excluding *Corse, Moselle, Bas-Rhin* and *Haut-Rhin*), according to the population of women. Population variations between the average population in 1939 and the recorded population in 1946. Migratory movement is the difference between the whole variation and the intrinsic growth. Intrinsic growth is the difference between births and deaths.

the total deaths in order to get the "shared mortality". Using both estimated populations and total flows of departmental births and deaths, I have calculated the annual internal migration flows during the Second World War, which had never been done before. These computations are all the more interesting as the differences in population variations between 1939 and 1946 are significant (between -20% and + 20% depending on the *département*) and cannot be explained by the differences in intrinsic growth (between -5% and +2%).

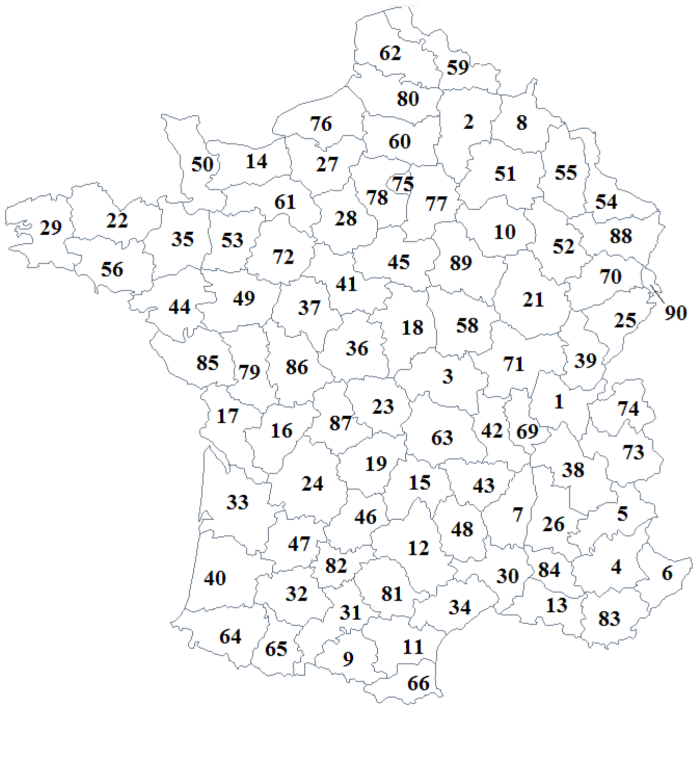
The annual monitoring of migratory flows has first revealed the extent of the May-June 1940 Exodus in France. Populations near the eastern borders, then in a further South as the German army advanced, fled to the South-West. 40% of *Ardennes*' population left between 1939 and 1940, as well as 10% of the Alpine *départements* one. In the northeastern *départements,* the population took a long time to come back, since the "northeastern line" prevented the return of refugees. This return was done later, with the easing of controls along this line. From 1941, the largest population movements moved from the North and North-West to the South, beyond the new demarcation line established by the German. While in the conventional wisdom this fictional border has been experienced as a tear, it is interesting to note that in fact the populations were able to cross it massively. In 1942 and 1943, the Mediterranean coastal *départements* hosted the equivalent of 20% of their 1939 populations, with maximum values reached in *Var* and *Alpes-Maritimes* at around

30%. From 1944, the refugee populations moved inland, fleeing the coasts by which the "*Libération*" arrived. Refugees in the Mediterranean area spread to neighboring rural *départements* such as *Lozère* or *Aveyron*. As such, *Basses-Alpes* was a haven for the South-East inhabitants afflicted by bombings. The situation gradually normalized between 1945 and 1946: it was "the backward Exodus". The southern *départements* that received the most of refugees saw them gradually moving back to the North and the North-West. At the end of the war, the French spatial demography still beared the scars of war: the refugees from *départements* north of an arc connecting *Loire-Inférieure* and *Doubs* via *Somme* did not all come back. The demographic weight of these missing women was between 5 and 15% of the 1939 population, with maximum values reached in *Meuse* or *Bretagne*.

This study, particularly innovative since it has used the quantitative information available through censuses and vital statistics, may open the way for further works on this historical event. In the first place, the study of males displaced would be challenging but quite interesting : one could quantify the gradual return of prisoners from Germany, the reverse flow of individuals belonging to the *Service du Travail Obligatoire* and maybe the migrations due to the *résistance*. Indeed, some *départements* were at the forefront of resistance: for example Drôme and Isère with the *Vercors*, but more generally the whole Alps and the *Cévennes*. This work would require both military deaths and deaths in deportation for each *département* and each year. The study of migratory movements according to age is also promising: it is likely that the elderly were less involved in these internal migrations because of their weakness, but no study has been able to quantify this phenomenon. Finally, other historical events during which migrations were intense could be studied, in France or elsewhere. As such, the *Pieds-Noirs* Exodus coming from Algeria in the 1960s is an emblematic one. The settlement's geography of these populations could be refined by using the method developed here, and would shed new light on this event, in line with Hunt (1992).

4.5 Appendices

4.5.1 Map of the 86 French *Départements*



1 Ain
2 Aisne
3 Allier
4 Alpes (Basses)
5 Alpes (Hautes)
6 Alpes Maritimes
7 Ardèche
8 Ardennes
9 Ariège
10 Aube
11 Aude
12 Aveyron
13 Bouches du Rhone
14 Calvados
15 Cantal
16 Charente
17 Charente Maritime
18 Cher
19 Corrèze

21 Côte d'Or
22 Côtes du Nord
23 Creuse
24 Dordogne
25 Doubs
26 Drôme
27 Eure
28 Eure et Loir
29 Finistère
30 Gard
31 Garonne (Haute)
32 Gers
33 Gironde
34 Hérault
35 Ille et Vilaine
36 Indre
37 Indre et Loire
38 Isère
39 Jura
40 Landes
41 Loir et Cher
42 Loire
43 Loire (Haute)
44 Loire Inférieure
45 Loiret
46 Lot
47 Lot et Garonne
48 Lozère
49 Maine et Loire
50 Manche
51 Marne
52 Marne (Haute)
53 Mayenne
54 Meurthe et Moselle
55 Meuse
56 Morbihan

58 Nièvre
59 Nord
60 Oise
61 Orne
62 Pas de Calais
63 Puy de Dome
64 Pyrénées (Basses)
65 Pyrénées (Hautes)
66 Pyrénées Orientales

69 Rhône
70 Saône (Haute)
71 Saône et Loire
72 Sarthe
73 Savoie
74 Savoie (Haute)
75 Seine
76 Seine Inférieure
77 Seine et Marne
78 Seine et Oise
79 Sèvres (Deux)
80 Somme
81 Tarn
82 Tarn et Garonne
83 Var
84 Vaucluse
85 Vendée
86 Vienne
87 Vienne (Haute)
88 Vosges
89 Yonne
90 Territoire de Belfort

The numbers are those used in Bonnet (2018)'s database. Since data are missing in the vital statistics, *Corse* (20), *Moselle* (57), *Bas-Rhin* (67) and *Haut-Rhin* (68) are not included in this paper.

4.5.2 Cause of Death Classification

Nomenclature	
1 to 14	Infectious and Parasitic Diseases
15	Cancer and other Malignant Neoplasms
16	Non-malignant neoplasms
17 to 20	General Diseases and Chronic Poisoning
21 to 23	Diseases of the Nervous System
24 and 25	Diseases of the Circulatory System
26 to 28	Diseases of the Respiratory System, without tuberculosis
29 to 34	Diseases of the Genetourinary System
35 to 36	Pregnancy, Childbirth and Puerperium
37	Skin, Cellular Tissue, Bones and Organs of Locomation Diseases
38	Débilité congénitale, vices de la conformation congénitaux, prématurés
39	Senility
40	Intentional Self-harm (*Suicide*)
41	Assault (*Homicide*)
42	Violent or Accidental Deaths (without Intentional Self-harm and Assault)
43	Causes non spécifiées ou mal définies

The nomenclature used to classify deaths according to their cause had 43 categories between 1936 and 1943. From 1944, the nomenclature added a 44^{th} category to take into account the deaths of car accidents. In this study, I have isolated deaths in the 42^{th} category (43^{th} after 1943), which correspond to what I call "asymmetrical mortality". They are mainly deaths due to bombings on the national territory, whose share in total deaths is 2.5 times higher in 1940 (6 times higher in 1944) compared to the pre-war level (1936).

4.5.3 Evolutions of Departmental Populations by Component, 1939–1946

	Total	Int. Growth	Migratory Mov.
Ain	-2.5%	-0.4%	-2.1%
Aisne	-4.2%	0.8%	-4.9%
Allier	7.6%	-2.0%	9.6%
Alpes (Basses)	-6.1%	-1.7%	-4.4%
Alpes (Hautes)	-0.1%	1.5%	-1.7%
Alpes Maritimes	-11.8%	-2.3%	-9.5%
Ardeche	-3.4%	-2.0%	-1.4%
Ardennes	-3.0%	1.3%	-4.3%
Ariege	-3.7%	-3.7%	0.0%
Aube	2.2%	-0.4%	2.6%
Aude	-2.6%	-2.7%	0.1%
Aveyron	3.9%	-1.1%	5.1%
Bouches du Rhone	-19.2%	-0.9%	-18.2%
Calvados	-8.9%	0.2%	-9.1%
Cantal	-2.2%	0.2%	-2.3%
Charente	-8.7%	-1.2%	-7.5%
Charente Maritime	-1.9%	-0.5%	-1.4%
Cher	-0.8%	-2.8%	2.0%
Correze	3.0%	-1.5%	4.5%
Cote d'Or	9.2%	-0.7%	9.9%
Cotes du Nord	-6.0%	0.1%	-6.1%
Creuse	-4.9%	-4.2%	-0.7%
Dordogne	-3.0%	-1.5%	-1.5%
Doubs	-3.2%	2.6%	-5.8%
Drome	6.6%	-1.2%	7.8%
Eure	1.7%	0.4%	1.3%
Eure et Loir	-0.7%	-0.5%	-0.2%
Finistere	-3.9%	1.1%	-5.0%
Gard	-2.9%	-1.9%	-1.0%
Garonne (Haute)	10.1%	-2.0%	12.1%
Gers	7.3%	-2.7%	10.1%
Gironde	-0.8%	-2.6%	1.7%
Herault	-7.6%	-2.1%	-5.5%
Ille et Vilaine	-6.4%	0.1%	-6.5%
Indre	-2.5%	0.2%	-2.7%
Indre et Loire	7.1%	-0.2%	7.2%
Isere	5.3%	-0.6%	5.9%
Jura	7.6%	0.6%	7.1%
Landes	3.1%	-2.0%	5.1%
Loir et Cher	-2.0%	-1.0%	-0.9%
Loire	2.8%	-0.5%	3.3%
Loire (Haute)	-10.5%	-3.2%	-7.3%
Loire Inferieure	-14.0%	-0.3%	-13.7%
Loiret	-4.1%	-1.3%	-2.8%
Lot	1.1%	-4.6%	5.7%
Lot et Garonne	13.7%	-1.2%	14.9%
Lozere	1.0%	-0.1%	1.0%
Maine et Loire	-3.6%	-0.4%	-3.2%
Manche	-7.1%	0.5%	-7.6%
Marne	-4.2%	1.0%	-5.1%
Marne (Haute)	-3.5%	0.1%	-3.6%
Mayenne	-4.3%	0.4%	-4.6%
Meurthe et Moselle	-2.8%	1.0%	-3.8%
Meuse	-8.7%	0.2%	-9.0%
Morbihan	-12.5%	0.4%	-12.9%
Nievre	-2.9%	-3.6%	0.7%
Nord	-2.3%	0.0%	-2.3%
Oise	-0.3%	-0.5%	0.1%
Orne	-10.3%	-0.7%	-9.6%
Pas de Calais	1.1%	2.8%	-1.7%
Puy de Dome	3.8%	-2.6%	6.4%
Pyrenees (Basses)	4.6%	-1.2%	5.9%
Pyrenees (Hautes)	10.8%	-2.2%	13.1%
Pyrenees Orientales	-13.6%	-1.7%	-11.9%
Rhone	4.9%	-0.7%	5.6%
Saone (Haute)	-6.2%	-1.4%	-4.7%
Saone et Loire	0.7%	-1.3%	2.1%
Sarthe	1.5%	1.1%	0.4%
Savoie	3.2%	0.9%	2.3%
Savoie (Haute)	12.6%	2.5%	10.1%
Seine	8.1%	-1.3%	9.4%
Seine Inferieure	-9.8%	0.4%	-10.2%
Seine et Marne	1.6%	-1.7%	3.3%
Seine et Oise	-5.2%	-2.5%	-2.7%
Sevres (Deux)	4.3%	1.1%	3.2%
Somme	-4.2%	0.0%	-4.2%
Tarn	-1.1%	-1.6%	0.5%
Tarn et Garonne	6.3%	-1.9%	8.2%
Var	-8.8%	-1.4%	-7.4%
Vaucluse	-2.8%	-1.3%	-1.5%
Vendee	-7.1%	0.7%	-7.8%
Vienne	-0.2%	0.3%	-0.4%
Vienne (Haute)	-10.8%	-1.8%	-9.0%
Vosges	-11.1%	-0.9%	-10.2%
Yonne	-0.9%	-4.2%	3.3%
Belfort	-0.7%	0.8%	-1.5%

4.5.4 Yearly Variations of Population due to Migratory Movement

	1939–1940	1940–1941	1941–1942	1942–1943	1943–1944	1944–1945	1945–1946
Ain	-7%	5%	-2%	1%	4%	-7%	3%
Aisne	-15%	-12%	9%	1%	2%	3%	7%
Allier	16%	-11%	-1%	2%	8%	-3%	-1%
Alpes (Basses)	-10%	4%	1%	13%	25%	-26%	-12%
Alpes (Hautes)	-16%	13%	-12%	24%	3%	-17%	3%
Alpes Maritimes	-8%	25%	1%	7%	-15%	-17%	-1%
Ardeche	-3%	-1%	6%	3%	5%	-9%	-3%
Ardennes	-42%	-21%	20%	26%	5%	-4%	11%
Ariege	-1%	2%	-8%	8%	10%	-8%	-3%
Aube	18%	-18%	-2%	-2%	7%	4%	-4%
Aude	2%	1%	4%	4%	6%	-14%	-4%
Aveyron	14%	-15%	2%	4%	11%	-13%	2%
Bouches du Rhone	-8%	22%	4%	-3%	-14%	-15%	-5%
Calvados	7%	-10%	-8%	-2%	6%	-8%	7%
Cantal	0%	-3%	-12%	6%	12%	-3%	-3%
Charente	8%	-20%	-2%	4%	-9%	19%	-8%
Charente Maritime	8%	-12%	-10%	5%	-1%	1%	6%
Cher	13%	-6%	-7%	-3%	6%	2%	-3%
Correze	16%	-7%	-7%	0%	2%	1%	0%
Cote d'Or	6%	-5%	-1%	4%	10%	-4%	0%
Cotes du Nord	-2%	-9%	-6%	0%	0%	11%	0%
Creuse	8%	-13%	-2%	1%	10%	-5%	1%
Dordogne	10%	-9%	2%	-5%	5%	-3%	-2%
Doubs	-13%	0%	4%	2%	3%	-6%	4%
Drome	2%	-1%	1%	5%	13%	-7%	-5%
Eure	6%	-13%	7%	-3%	6%	-6%	5%
Eure et Loir	9%	-13%	0%	3%	2%	0%	-2%
Finistere	0%	-10%	-7%	-1%	1%	4%	7%
Gard	2%	6%	2%	11%	-2%	-13%	-8%
Garonne (Haute)	1%	12%	2%	-2%	-5%	-1%	4%
Gers	8%	-3%	-8%	6%	1%	1%	5%
Gironde	8%	-5%	11%	-13%	-5%	6%	0%
Herault	5%	8%	6%	-7%	2%	-17%	-1%
Ille et Vilaine	-2%	-8%	-4%	-2%	-4%	18%	-5%
Indre	6%	-12%	-12%	5%	2%	12%	-5%
Indre et Loire	10%	-6%	-2%	2%	7%	9%	-13%
Isere	-4%	9%	10%	-1%	6%	-3%	-11%
Jura	-5%	4%	1%	-1%	-2%	0%	10%

	1939–1940	1940–1941	1941–1942	1942–1943	1943–1944	1944–1945	1945–1946
Landes	2%	-9%	-4%	3%	-2%	11%	5%
Loir et Cher	8%	-14%	13%	-4%	-9%	6%	-3%
Loire	0%	6%	3%	4%	1%	-5%	-6%
Loire (Haute)	4%	-2%	-8%	7%	8%	-4%	-11%
Loire Inferieure	-8%	-2%	-8%	3%	-11%	15%	-3%
Loiret	16%	-23%	0%	1%	6%	2%	-3%
Lot	2%	-5%	3%	3%	11%	-5%	-3%
Lot et Garonne	14%	-9%	-2%	11%	7%	-5%	-1%
Lozere	10%	-21%	2%	10%	18%	-18%	1%
Maine et Loire	2%	-8%	0%	4%	7%	9%	-17%
Manche	-1%	-9%	-8%	4%	-3%	4%	7%
Marne	-1%	-12%	-1%	3%	3%	8%	-6%
Marne (Haute)	4%	-23%	-1%	3%	11%	2%	0%
Mayenne	5%	-15%	-9%	-1%	8%	4%	4%
Meurthe et Moselle	-16%	1%	5%	7%	3%	-7%	3%
Meuse	-10%	-25%	7%	14%	-1%	3%	3%
Morbihan	1%	-7%	-6%	-8%	-7%	9%	4%
Nievre	11%	-15%	1%	1%	12%	-2%	-7%
Nord	-3%	-7%	-3%	2%	-5%	4%	9%
Oise	2%	-3%	5%	-7%	1%	-4%	6%
Orne	-3%	-14%	-2%	0%	12%	-2%	0%
Pas de Calais	-1%	-12%	-1%	3%	-8%	6%	11%
Puy de Dome	12%	-8%	-5%	3%	4%	-4%	4%
Pyrenees (Basses)	12%	-7%	-2%	-2%	-3%	5%	4%
Pyrenees (Hautes)	9%	-9%	3%	6%	10%	-11%	4%
Pyrenees Orientales	-17%	10%	1%	1%	4%	-15%	3%
Rhone	3%	11%	9%	-5%	5%	-9%	-9%
Saone (Haute)	-9%	6%	6%	-1%	5%	-18%	7%
Saone et Loire	1%	-3%	1%	-2%	6%	-2%	0%
Sarthe	4%	-12%	-2%	5%	3%	15%	-12%
Savoie	-11%	14%	2%	-9%	16%	3%	-13%
Savoie (Haute)	-8%	7%	-2%	8%	5%	2%	-2%
Seine	-8%	23%	-3%	-4%	-1%	0%	3%
Seine Inferieure	-1%	-2%	4%	-16%	-7%	8%	3%
Seine et Marne	7%	-9%	3%	2%	-1%	-1%	2%
Seine et Oise	-5%	13%	0%	-7%	-7%	0%	4%
Sevres (Deux)	12%	-15%	-4%	1%	3%	13%	-8%
Somme	-5%	-11%	4%	1%	-5%	7%	6%
Tarn	-1%	-2%	3%	-3%	10%	-8%	2%
Tarn et Garonne	8%	-2%	-1%	-2%	10%	-7%	2%
Var	-5%	24%	12%	-1%	-24%	-8%	-6%
Vaucluse	-4%	11%	7%	6%	3%	-20%	-5%
Vendee	9%	-15%	-7%	7%	-4%	11%	-9%
Vienne	17%	-29%	5%	5%	-12%	28%	-13%
Vienne (Haute)	14%	-22%	4%	3%	9%	-6%	-12%
Vosges	-7%	-7%	0%	1%	13%	-13%	2%
Yonne	16%	-17%	-1%	-1%	7%	-3%	2%
Teritoire de Belfort	-7%	-4%	0%	5%	3%	1%	0%

Part II: Data and Methods

Chapter 5

Computations of French Lifetables by *Département*, 1901–2014

Abstract

Debates concerning the territorial divide in France are deep. To bring a contribution to this issue, I compute the departmental lifetables since 1901, for both men and women. In this paper, I present the raw data collected to do so, namely yearly births and deaths by age as well as population by age at each census carried out during the 20th century. I add statistics according to military mortality and mortality in deportation to cover the periods of the Two World Wars. I also present the methods I use to compute these lifetables, which come mainly from the Human Mortality Database protocol. I revise this protocol to take into account the specificities of French departmental data, mainly the few changes in French departmental boundaries, the underestimation of infant mortality and the lack of raw data homogeneity. This new database complements a still limited supply of long-term mortality statistics computed at local level.[1]

[1]This study received financial support from the ERC Grant "Demographic Uncertainty" led by Hippolyte d'Albis. The author would especially like to thank Magali Barbieri and Hippolyte d'Albis for their many advices, as well as the participants of the Working Group *"Valeurs Extrêmes et Longévité"* (University Paris VI), the participants of INED seminars *"Histoire et Populations"* and *"Mortalité, Santé, Epidémiologie"*, and the participants of ENS Paris-Saclay internal seminar.

5.1 Introduction

Life expectancy has risen sharply in France since the beginning of the 20th century. The lifetables calculated by Vallin and Meslé (2001) for the 19th and 20th centuries show that men life expectancy at birth was 33 in 1806, 44.5 in 1898, 60 in 1946 and 74.5 in 1997. This increase at the national level does not say anything about increases at the local level. As such, significant differences exist between the French *départements*. Barbieri (2013) worked on departmental mortality and showed that the life expectancy at birth of men for the period 2006–2008 was 74.4 years in *Nord*, compared to 79.7 years in *Hauts-de-Seine*, a difference of more than 5 years. This situation explains why the debate on the territorial divide according to health is important in France. Indeed, departmental differences can not be explained from a public policy point of view: the State has to reduce these inequalities. In order to inform public decision-makers in their choices, it is important to know the history of these departmental differences.

Consequently, I compute in this paper the yearly departmental lifetables by sex for all French metropolitain *départements* between 1901 and 2014. The computation of these lifetables is based on the exhaustive collection of population flows (deaths by age and sex, births by sex) and population stocks at each census (population by age and sex). I exploit a French unique characteristic: since 1789, this country is divided into around 100 geographical units of similar size, namely *départements*. This division has changed very little during two centuries, and the statistical centralizations have been carried out at this geographical level. Moreover, in order to take into account the two World Wars that affected France between 1914–1918 and 1939–1945, I have collected in two original sources the military deaths by age during the two wars as well as the deaths in deportation by age and sex during the Second World War. With these lifetables, I get life expectancies and mortality rates at each age for more than 100 years. In addition, I get populations by age and sex at each January 1^{st}.

These lifetables at the subnational level complete a still incomplete literature. Bonneuil (1997) worked on departmental mortality in the 19th century: he computed women lifetables by five-year period and for five-year age groups. He followed Van de Walle (1974) who computed similar lifetables with a different methodology. These two authors have not studied in the same way men's mortality, because of strong fluctuations due to the wars which afflicted France at this time. From 1954 to 1999, Daguet (2006) grouped lifetables established at the departmental level, but only for the census years. Barbieri (2013) used in her study departmental lifetables calculated by INSEE for the period 1975–2008 . However, these data were provided exceptionally. Vallin and Meslé (2005) used departmental life expectancies for the period 1906–1954. However, both reconstruction methods and data have never been published. Lastly, various mortality indicators are available in official publications, namely *Statistique Annuelle du Mouvement de la Population*.[2] However, these indicators are relatively scarce: they relate only to infant mortality rates, or standardized mortality rates.

In addition, the lifetables I compute are based on a unified methodological protocol for the whole period 1901–2014, which is not the case of the papers previously cited. This methodological protocol is available in Wilwoth et al (2007). Many researchers are using this protocol to compute national

[2]See, for example, page 74 of this publication for the year 1938.

lifetables for a large number of countries. It is also used to compute lifetables at the local level in two OECD countries. The results according to Canadian provinces for the period 1921–2011 are available in the Canadian Human Mortality Database[3], and those according to the Japanese provinces since 1975 are available in the Japan Mortality Database.[4] This paper therefore complements a still limited supply of local mortality data freely available by adopting an internationally recognized protocol; this allows international comparisons without methodological bias.

The rest of this paper is organized as follows. In Section 2 I present the statistical sources used to compute departmental lifetables. The methods used are explained in Section 3 in which I distinguish the methods coming from the HMD protocol and the methods specific to this study. In Section 4 I illustrate some of the results available in this new database. Part 2 of this paper is the methodological appendix.

5.2 Sources

Computations of departmental lifetables requires two types of data: population movement (deaths and births domiciled), and population censuses. The deaths collected do not only concern civilian deaths: both military deaths during the two World Wars and deportation deaths between 1939 and 1945 have been included.

5.2.1 Deaths

Civilian deaths of each *département*, each sex and each year over the period 1901–2014 have been retrieved from the population movement statistics published by *Statistique Générale de la France* (SGF) and then by *Institut National de la Statistique et des Etudes Economiques* (INSEE). I have retrieved deaths by age group recorded in home *département*. Tables 5.7 and 5.8 in Appendix (Part 2, Section 5.6.9) provide sources in which raw statistics have been found. In addition, I have collected in Vallin and Meslé (2001) single-age and sex-specific civilian deaths at the national level for the same period.

I have retrieved deaths during the two World Wars from Defense Ministry's website.[5] They are available by year of birth at the departmental level, and by year of birth and year of death at the national level.

Individuals who died during deportation in the Second World War are not included in the civilian population movement. However, they were nearly 100,000. I have decided to include them in my statistics, using data from Memorialgenweb Website.[6] This database records deportees who left France and died in deportation published in the *Journal Officiel*, by *département* of birth if they were born in France, and by country of birth otherwise. Table 5.1 presents figures of the foreign-born deportees by country of birth. One can see that the Poles were the most numerous. Although this

[3]Computed by researchers in "*Université de Montréal*", www.demo.umontreal.ca/chmd/.
[4]Computed by researchers at the National Institute of Population and Social Security Research, http://www.ipss.go.jp/p-toukei/JMD/index-en.asp.
[5]http://www.memoiredeshommes.sga.defense.gouv.fr/
[6]http://www.memorialgenweb.org/memorial3/deportes/index.php, forwarded on March 7th, 2016

database is not exhaustive, the large number of observations provides a sample close to the total of deaths in deportation.

Table 5.1: Summary of foreign-born deportees by nationality

Country	Deportees	In % of foreign-born deportees
Pologne	13,599	40.46%
Spain	5,075	15.10%
Russia	2,741	8.16%
Germany	2,425	7.21%
Romania	1,861	5.54%
Turkey	1,511	4.50%
Algeria	1,050	3.12%
Greece	939	2.79%
Italia	535	1.59%
Ukraine	534	1.59%

Notes: Deportees by country of birth in the Memorialgenweb's database.

5.2.2 Births

I have retrieved births by year, sex and mother's home *département* for the period 1901–2014. I have also recovered stillbirths by mother's home *département* and year (both males and females). Finally, I have retrieved births by year, sex and mother's home *département* for the period 1853-1900.[7]

5.2.3 Censuses

Finally, I have collected populations by birth year, home *département* and sex for each census of the period 1901–1962 from hard-copy publications of SGF and INSEE. For the period 1968–2014, these statistics have been found in on-line sources. These data are not available for each year because censuses were held at varying intervals. Between 1901 and 2014, censuses were made in 1901, 1906, 1911, 1921, 1926, 1931, 1936, 1946, 1954, 1962, 1968, 1975, 1982, 1990, 1999, 2008, 2013 and 2014.[8]

5.3 Methods

The protocol I use to compute departmental lifetables is largely inspired by the one of the Human Mortality Database (HMD). This database gathers all national lifetables computed using these methods. However, since my database is specific both for the small numbers in each *département* and the time period chosen (including the two World Wars), I have added specific methods.

[7]Tables 5.9, 5.10 and 5.11 in Appendix (Part 2, Section 5.6.9) give sources in which raw statistics have been found.
[8]Table 5.12 in Appendix (Part 2, Section 5.6.9) gives sources in which raw statistics have been found.

5.3.1 HMD Protocol Methods

5.3.1.1 Raw Data Adjustments

Raw data adjustments according to deaths are the main issue since they are aggregated into five-year age groups until 1967 and by single age between 1968 and 2014. To get a 1×1 format (single age, year of death) for the deaths between 1901 and 1967, I distribute deaths at unknown age among age groups, and adjust the curve of cumulative deaths by cubic splines. Cubic Spline is a semi-parametric estimation method which joins the points of a cumulative distribution by third degree polynomials. Let $Y(x) = \sum_{u=0}^{x-1} D_u$ be the cumulative number of deaths up to age x. $Y(x)$ is known for a limited collection of ages including 1, 5, 10... etc from the raw data. I know $Y(x)$ for both the highest age in the distribution (80, 90 or 100) and the age above which no further deaths are observed, set at 105. Equation (5.3.1) fits a cubic spline by using these values (the indicator function $I(.)$ equals one if the logical statement within parentheses is true and zero otherwise):

$$Y(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \beta_1 (x - k_1) I(x > k_1) + ... + \beta_n (x - k_n) I(x > k_n). \qquad (5.3.1)$$

I have to estimate the vector $(\alpha_0; \alpha_1; \alpha_2; \alpha_3; \beta_1; ...; \beta_n)$ which contains $n + 4$ coefficients, but I only know $n + 2$ values of $Y(x)$, and therefore $n + 2$ constraints. Two further constraints must be introduced to identify the model. First I assume that there is no death at the upper bound, namely 105. Second I assume that deaths observed between 1 and 5-year-old occured between 1 and 2 year-old. $\hat{Y}(x)$ are calculated for all ages, for each *département*, sex and year. Deaths at age x are found as follows:

$$\hat{D}(x) = \hat{Y}(x + 1) - \hat{Y}(x).$$

Negative death counts may occur when the deaths in five-year age groups are extremely low.[9] The method is to set zero-deaths in age groups where negative counts occur. To balance this, deaths in the adjacent age groups are reduced pro-rata their number of deaths. If D_{neg} is the sum of negative death counts for an observation, D_s^* the deaths at age s after allocation of negative death counts, D_s the estimated deaths at age s before allocation of negative death counts, x_1 and x_2 the lower and higher limits of the interval in which the negative death counts are observed, then:

$$\begin{cases} D_s^* = 0 & \text{for } s \in [x_1, x_2]\,, \\ D_s^* = D_{neg} \times \frac{D_s}{\sum_{i \in \Omega_1} D_i} & \text{for } s \in [x_1 - 5, x_1] \cup [x_2, x_2 + 5]\,, \\ D_s^* = D_s & \text{otherwise.} \end{cases} \qquad (5.3.2)$$

Deaths estimated by cubic spline are too imprecise to be used at advanced ages: open-age interval of deaths is too low (see Tables 5.7 and 5.8, Column 5). These deaths are adjusted by means of the Kannisto model, which assumes a survival curve of logistic form, with a zero-asymptote for very old

[9]This usually happens at around age 30. I count only seven of these occurrences, but they need to be adjusted so as not to have negative mortality rates later.

ages. I use this method for deaths beyond the open-age interval – different according to the periods, I keep a maximum of 95 so that estimates are not hindered by too small figures – and rely on the deaths observed for ages 10 years below this limit. Thus, if the open-age interval begins at age 90, I use the ages 80–89. Formally, I compute a fictitious survival curve $S(80+x)$:

$$S(80+x) = \frac{\sum_{u=80+x}^{105} D_u}{\sum_{u=80}^{105} D_u} \quad \text{for } x = 0,\ 1,\ 2,\ ...,\ 9. \tag{5.3.3}$$

This survival function conditional on reaching age 80 may be seen as tracking a "synthetic extinct cohort", since it is based on annual deaths and not on deaths in the cohort itself. Assuming that this fictitious cohort displays survival probabilities that can be fitted by the Kannisto model, the survival function $s(x)$ is:

$$s(x) = \left(\frac{1+a}{1+ae^{b(x-80)}}\right)^{1/b}. \tag{5.3.4}$$

with estimated values for a and b, I compute $\hat{s}(x)$ et $d(x) = \hat{s}(x) - \hat{s}(x+1)$. Finally, I obtain deaths at each age:

$$D(x) = \sum_{u=90}^{105} D_u \times \frac{d(x)}{\hat{s}(90)}. \tag{5.3.5}$$

I finally proceed to a uniform adjustment so that the sum of the departmental deaths for each age, year, and sex corresponds to the national data.

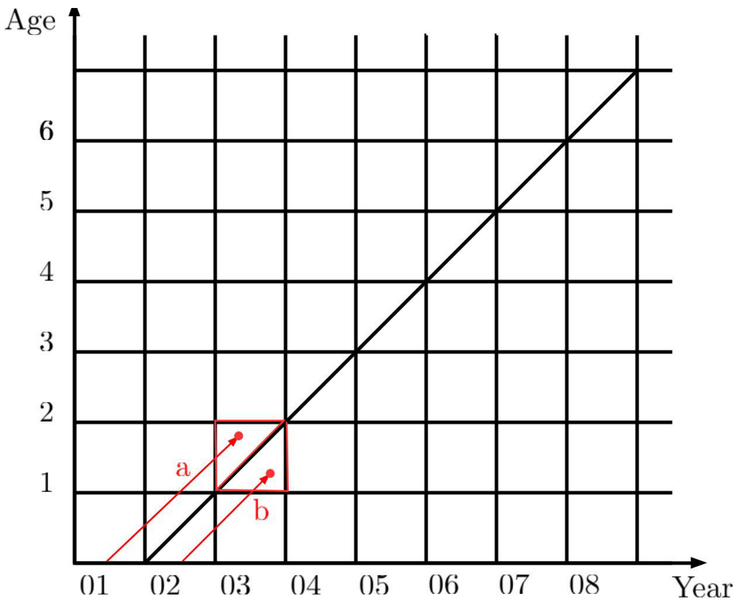
For censuses, raw data are generally available for groups of five-year of births. I use the Cubic Splines method in the same way to estimate populations according to their year of birth.

5.3.1.2 Splitting Deaths into Lexis Triangles

Figure 5.3.1 presents deaths by year and age. They may be split into two triangles for a single year, known as Lexis triangles. For individuals who died between ages 1 and 2 in 1903, one may distinguish two kinds of deaths. The first who died between ages 1 and 2 in 1903, born in 1901 ("a" on Figure 5.3.1, upper triangle). The others who died between ages 1 and 2 in 1903, born in 1902 ("b" on Figure 5.3.1, lower triangle).

Overall, if the probability of death is equiprobable over time, one could think that the distribution of annual deaths by age for half in the lower triangle and the other half in the upper triangle would be sufficiant. This is not, for two main reasons. The first is that infant mortality, when high, is observed largely in the first days after birth, and must therefore be integrated into the lower triangle. The second concerns the relative size of cohorts, which also influences the distribution between triangles. When the flow of births varies greatly from one year to the next (e.g. during the two World Wars), the half-death distribution in the lower triangle is strongly biased. The HMD protocol sets a sex-specific equation allowing the distribution of deaths in Lexis triangles. This equation takes into account the relative size of two successive cohorts, age, some historical events (e.g Spanish influenzia), and the infant mortality rate. If we call x the age and t the year, these sex-specific equations are as follows (Equation (5.3.6) for women, Equation (5.3.7) for men):

Figure 5.3.1: An exemple of lexis diagram



$$\begin{aligned}\hat{\pi_d}(x,t) =& 0.4710 + \hat{\alpha_F} + 0.7372\,[\pi_b(x,t) - 0.5] \\ &+ 0.1025\,I_{t=1918} - 0.0237\,I_{t=1919} \\ &- 0.0112\,logIMR(t) - 0.0688\,logIMR(t)\,I_{x=0} + 0.0268\,logIMR(t)\,I_{x=1} \\ &+ 0.1526\,[logIMR(t) - log(0.01)]\,I_{x=0}I_{IMR(t)<0.01}\end{aligned}; \tag{5.3.6}$$

$$\begin{aligned}\hat{\pi_d}(x,t) =& 0.4836 + \hat{\alpha_H} + 0.6992\,[\pi_b(x,t) - 0.5] \\ &+ 0.0728\,I_{t=1918} - 0.0352\,I_{t=1919} \\ &- 0.0088\,logIMR(t) - 0.0745\,logIMR(t)\,I_{x=0} + 0.0259\,logIMR(t)\,I_{x=1} \\ &+ 0.1673\,[logIMR(t) - log(0.01)]\,I_{x=0}I_{IMR(t)<0.01}\end{aligned}. \tag{5.3.7}$$

$\hat{\pi_d}(x,t)$ is defined as the proportion of death of a given year and age allocated in the lower triangle. α_F and α_H are age-specific values coming from the HMD protocol.

$\pi_b(x,t)$ is defined as the ratio of births between two successive cohorts and calculated only once for both sexes:

$$\pi_b(x,t) = \frac{B(t-x)}{B(t-x) + B(t-x-1)}. \tag{5.3.8}$$

Long historical series are required to calculate this ratio for all the cohorts tracked between 1901 and 2014. One can take individuals aged 80 in 1901 as an example. To calculate this ratio one needs birth in 1820 and 1821. I was unable to do so: my birth records only go back to 1853. For earlier years I assume that births before 1853 were equal to births in 1853 and use a birth ratio of 0.5.

$IMR(t)$, the same for both sexes, is calculated as follows:

$$IMR(t) = \frac{D\,(0,t)}{\frac{1}{3}B(t-1) + \frac{2}{3}B(t)}. \tag{5.3.9}$$

If births are not available for one of the two years, $IMR(t)$ is calculated as follows[10]:

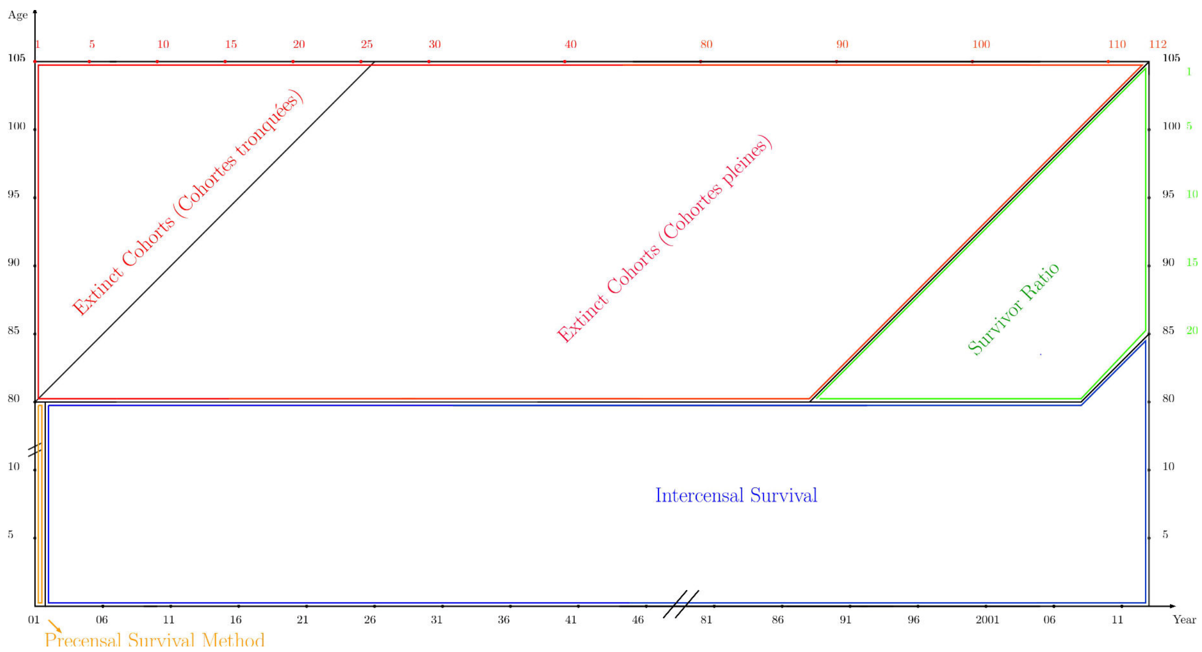
$$IMR(t) = \frac{D\,(0,t)}{B(t^*)}, \tag{5.3.10}$$

with t^* the year for which births are available.[11]

5.3.1.3 Computations of Populations by Age at 1^{st} January of each Year

To calculate the mortality rates required for lifetables, I need populations by age at 1^{st} January for each year from 1901 to 2014. I get populations by age in 2014 from official statistics so I may calculate populations by age for the 1901–2013 period.[12] Figure 5.3.2 reveals the four methods used for various periods and ages. Section 5.6.1 of the Appendix precisely specifies each method used.

Figure 5.3.2: Methods for computations of population at 1^{st} January



Notes: Methods used to compute populations by age at each 1^{st} January. For more details, see Section 5.6.1 of Appendix.

The "Intercensal Survival" method is used to estimate the population under age 80 from 1902 to 2013. Starting from one census (say, 1901) the population by age at the following census (1906) is

[10]When $IMR(t)$ is equal to zero because of no infant deaths, I assume a 0,00000001 IMR value so that $logIMR(t)$ can be calculated.

[11]I obtain proportions of deaths in the lower triangle greater than 1 for 28 female orbservations and 30 for male observations, all in 1918 or 1919 and for deaths under age 1. This is due to the Spanish influenza epidemic, the high infant mortality rate and the size differences between the cohorts born in 1918 and 1919. To tackle this issue, the death proportions in the lower triangle are set at 1, leading to zero death in the upper triangle for these observations.

[12]I do not need to calculate populations by age in 2014 since these data are available in raw statistics provided by INSEE

estimated by subtracting from the population by age in 1901 the deaths that occurred from 1901 to 1906. The difference between estimated and recorded populations in 1906, due to measurement errors and migrations, is then attributed to the intercensal population figures.

Second, the "Precensal Survival" method is used to estimate the population under age 80 in 1901. Since "Intercensal Survival" estimates the population under age 80 on 1^{st} January of each intercensal year, I cannot compute population by age on 1^{st} January the year of the first census. To do so, I use the population on the day of the 1901 census and add the deaths occurring between 1^{st} January 1901 and the census day. Since there is no second census available as with "Intercensal Survival " method, I cannot correct for migrations and errors: I assumed that in so short a period these are minimal.

With the "Extinct Cohorts" method I can estimate the population aged 80 and over born in the cohorts that died between 1901 and 2013. I assume that migrations after age 80 are small; I compute the population of a cohort still alive by summing its future observed deaths.

Finally I estimate the population aged 85 and over in 2014 with the "Survival Ratio" method. I assume that the survival ratio between two ages for the extinct cohorts can be applied to the still living cohorts in oder to estimate their size at the last census. The 85-and-over estimated are then adjusted by the 85-and-over recorded in 2013. After this adjustment I compute the size of the intermediate populations located in the green quadrilateral by substracting step by step the observed deaths.

5.3.1.4 Adjustment of Computed Mortality Rates

I can compute departmental mortality rates by age and sex with deaths in Lexis triangles and populations at each 1^{st} January. Mortality rates are the ratio between the number of deaths and the number of individuals exposed to the risk[13]:

$$M_{xt} = \frac{D_{xt}}{E_{xt}} = \frac{D_L(x,t) + D_U(x,t)}{\frac{1}{2}\left[P(x,t) + P(x,t+1)\right] + \frac{1}{6}\left[D_L(x,t) - D_U(x,t)\right]}. \tag{5.3.11}$$

Note that I do not calculate populations for 2015, although these are needed for 2014. To estimate mortality rates for that year, I assume that the population at each age in 2015 is equal to that in 2014, and the formula becomes:

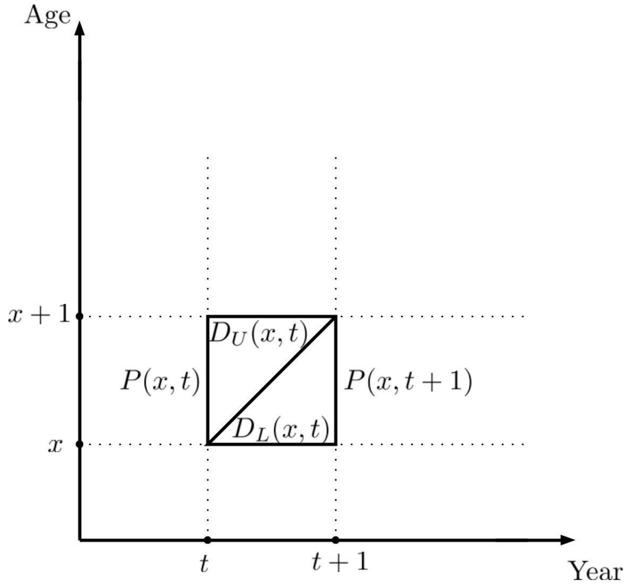
$$M_{x2014} = \frac{D_{x2014}}{E_{x2014}} = \frac{D_L(x,2014) + D_U(x,2014)}{P(x,2014) + \frac{1}{6}\left[D_L(x,2014) - D_U(x,2014)\right]}. \tag{5.3.12}$$

Figure 5.3.3 presents the set of data needed to compute mortality rates.

These rates are not used directly to calculate lifetables. I smooth mortality rates beyond age 90 in order to avoid erratic fluctuations due to small numbers of deaths and population at risk. The instantaneous probability of dying over age 80 in the Kannisto model can be expressed as follows (with a and $b \geq 0$):

[13] For the explanation of the presence of the difference between the two Lexis triangles at the denominator, please see HMD Protocol, Appendix E.

Figure 5.3.3: Mortality rates computations



$$\mu_x(a,b) = \frac{ae^{b(x-80)}}{1+ae^{b(x-80)}}. \tag{5.3.13}$$

Mortality rates estimated with the Kannisto model $M_x(a,b)$ are:

$$M_x(a,b) = \mu_{x+0,5}(a,b). \tag{5.3.14}$$

If $D_x \sim Poisson\left(E_x\mu_{x+0,5}(a,b)\right)$, then parameters a and b may be calculated by minimizing the following function:

$$-\,logL(a,b) = \sum_{x=80}^{105}\left[D_x log\mu_{x+0,5}(a,b) - E_x\mu_{x+0,5}(a,b)\right]. \tag{5.3.15}$$

I can calculate $\hat{M}_x(\hat{a},\hat{b})$ for all ages above 90, with estimated parameters $\left(\hat{a},\hat{b}\right)$. I assume that the population's mortality rates are equal to the mortality rates in the survival tables (m_x):

$$\begin{cases} m_x = M_x & x \in [0,89] \\ m_x = \hat{M}_x & x \in [90,105] \end{cases}. \tag{5.3.16}$$

To convert the survival table mortality rates into probabilities of dying, one must define a_x, the mean number of years lived by people dying between ages x and $x+1$. I assume that deaths are uniformly distributed at each age:

$$\begin{cases} a_x = 1/2 & x \in [1,104] \\ a_x = \frac{1}{m_{105}^{\infty}} & x = 105+ \end{cases}. \tag{5.3.17}$$

For age 0, I follow Preston (2001), who refers on Coale and Demeny (1983)'s lifetables. Thus:

$$\begin{cases} m_0 \geq 0.107 & \begin{cases} a_0 = 0,350 & \text{for women,} \\ a_0 = 0,330 & \text{for men,} \end{cases} \\ m_0 < 0.107 & \begin{cases} a_0 = 0,053 + 2.800 & \text{for women,} \\ a_0 = 0,045 + 2.684 & \text{for men.} \end{cases} \end{cases} \tag{5.3.18}$$

The probabilities of death may be calculated as follows:

$$\begin{cases} q_x = \frac{m_x}{1+(1+a_x)m_x} & x \in [0, 104] \\ q_x = 1 & x = 105+ \end{cases}. \tag{5.3.19}$$

With values of q_x, I can compute each of the lifetable values, for each age: the number of survivors (l_x), the number of deaths (d_x), and the life expectancies (e_x). Two lifetables are estimated: complete in format (1×1) i. e. for each age and each year, and in the format (1×5) i. e. for each age and each group of 5 years. For the sake of readability, lifetables in the (1×5) and (5×5) formats are also estimated. So I get values for age groups $[0, 1[$, $[1, 5[$, $[5, 10[$, $[10, 15[$ ... etc until ages 105 and over. Section 5.6.2 of the Appendix reviews the computations made to estimate each of the outstanding lifetable values in each specification.

5.3.2 Specific Departmental Methods

The methods presented previously come from the Human Mortality Database protocol. However, they are too general to be applied without correction to the case of French *départements* during the 20th century. These corrections are due to three main issues: the quality of the raw data, the two World Wars, and the territorial changes in my departmental classification.

5.3.2.1 Specific Methods Due to Data Quality

I include false stillbirths in births and deaths before first birthday as Vallin and Meslé (2001) did for the national lifetables. In their work they explained that before 1993, a child born alive who died before the official statement of birth was considered to be stillborn, which distorts both deaths before first birthday and births. To reduce this bias I have retrieved from official publications the false stillbirths by sex at the national level (Vallin and Meslé, 2001) and I have distributed them among *départements* pro rata of stillbirths. I added them to deaths before first birthday and births.

Moreover, the data retrieved from censuses are not of identical quality so I make some adjustments. The first is to distribute individuals of unknown year of birth pro rata of the numbers in known year of birth age groups. Although this do not present any problem for most censuses, this is not true for the 1901 one, when these numbers were included in those of the open-age interval. The second is to split the open-age interval 80-year-old and over in the 1906, 1921, 1926, 1931, 1936 and 1946 censuses. This open-age interval occurs too soon and generated some negative population figures. I split it in two age groupes: ages 80 to 84 and 85 and over. For these first two adjustments I use the 1911 census particularly detailled. Moreover, younger age groups did not always use the same variables: sometimes year of birth, sometimes age. I use a linear interpolation to compute figures per year of birth. Section 5.6.3 of the Appendix presents in a more detailled manner these threee adjustments.

5.3.2.2 Specific Methods Due to the Two World Wars

The two World Wars had significant demographic effects both at national and departmental level. The first is due to internal migrations caused by the conflict and the France's division into occupied and unoccupied zones in 1940. The raw statistics give no direct indication for this question. The second concerns the heavy military losses, which had to be included in death statistics. On this particular point, this study is the first to integrate military and deportation deaths into lifetables at subnational level.

Ideally, the statistics of military deaths should be available according to the age and the year of the soldier's death, as well as his home *département* before the war. Since the sources used are incomplete, I couple two different matrices. The first provides the total of deaths by *département* and year of birth. It comes from the Defense Ministry's database, which lists all the *"Morts pour la France"* (MPLF) of the two wars. The second provides the total of deaths at the national level by year of death and year of birth. It mobilizes the crowd-based indexing on the *Mémoire des Hommes* website: each individual, using his personal research on a specific soldier, inform both his year of death and his year of birth. This work has been done for just over 20% of total deaths. I wonder if this sample is representative of the distribution by year of death. For that, I use Pedroncini (1992)'s work: it gives total military deaths by year of death. Table 5.2 shows these distributions according to both sources. Even if discrepancies exist, I can use the sample coming from *Mémoire des Hommes*. Data by year of birth and year of death are therefore extracted from the Defense Ministry's database.

Table 5.2: Distribution by year of death of soldiers

	Year	1914	1915	1916	1917	1918	Total
Mémoire des hommes	Deaths	75,403	82,878	50,933	34,436	52,459	296,109
	% of the total	25.46%	27.99%	17.20%	11.63%	17.72%	100%
Pedroncini (1992)	Deaths	301,000	349,000	252,000	164,000	235,000	1,301,000
	% of the total	23.14%	26.83%	19.37%	12.61%	18.06%	100%

By cross-referencing these two matrices, I get a matrix giving total deaths by *département*, year of birth and year of death. I assume that there is little variation between *départements* in the year of death according to the cohort.

This distribution of deaths is then adjusted by the total of deaths as estimated by researchers at national level, so as to verify the overall consistency of the various sources. Prost (2008) makes an inventory of the statistical estimates of deaths during the First World War. He used the Marin's report, followed by Hubert (1931) and Dupaquier (1988). Roure's report cited by Prost (2008) revealed 1,357,800 military casualties, taking into account deaths of foreigners. Hubert (1931) added 40,000 soldiers dead during the 6 months after the armistice as well as sailors. Table 5.3 summarizes these numbers. Regarding the 28,600 deaths that occurred 6 months after the armistice, I assume that they had been included in the 1919 deaths of the population movement and do not take them into account. With regard to the 75,700 deaths of soldiers coming from settlements and abroad, since these populations were not registered in 1911 in the French *départements* and were surely recorded in the civilian deaths of their home country, I do not keep them in the total. Finally, I obtain 1,304,400 deaths.

Table 5.3: Militarty deaths during the First World War

Source	Variable	Deaths
Roure	Total of French military deaths	1,282,100
	Total foreign-born and settlements	75,700
	Total Roure	1,357,800
Hubert	Deaths 6 months after armistice	28,600
	Sailors	11,400
	Final total	1,397,800

The principle is the same for the Second World War. The two matrices combined come from the Defense Ministry's database. The total of deaths I use is 200,000, in line with Lagrou et al. (2002). Section 5.6.4 of the Appendix reviews the departmental classification problems in the Defense Ministry website, as well as the cubic splines used to distribute departmental deaths by single year of birth for the two World Wars.

According to deportation during the Second World War, deportees are classified by birth place in the database, which is different from home place. I build cross-matrices between birth place and home place for the deportees born in France and those born abroad. For that purpose I use two raw materials. The first is the 1936 census for the foreign-born, which provides their distribution among *départements* in France. The second is the 1946 census for the French-born, which provides their distribution by birth place and home place at departmental level. Finally, I adjust these figures by the total of deportees estimated by researchers, namely 110,000, in line with Dupaquier (1988). Section 5.6.5 of the Appendix presents the computations of deportees by age, sex and home-*département.*

5.3.2.3 Specific Methods Due to Territorial Changes

The main advantage of the French *départements* is their stability since the beginning of the 19th century. However, there were some changes during the two last centuries, especially with regard to the eastern borders and the Paris region. To take this into account, some adjustments are necessary. In this study, I use a departmental classification with 97 *départements*: the 95 *départements* of the current metropolitan France (*Corse* counting as one), as well as the *Seine* and *Seine-et-Oise* in their pre-1968 boundaries. Territorial breakdowns are twofold in this study: either departmental boundaries changed because of a territorial reorganization, or the data are missing within the unified departmental classification that I use.

The departmental boundary changes are of two types for the period 1853–2014. The first concerns the pre-1901 period. *Savoie* and Nice's *Comté* were attached to France following the April, 22th and 23th, 1860 *plébiscite*. *Savoie* and *Haute-Savoie* were created ex nihilo on June 14th, 1860 while

Alpes-Maritimes was created by aggregating a part of *Var* (*Grasse*'s *canton*) to the *Comté*. Moreover, following the war against Prussia in 1870, *Meurthe* and *Moselle* in their old form disappeared to form *Moselle* and *Meurthe-et-Moselle.*[14] In addition, the *départements* boundaries of *Haut-Rhin*[15], *Bas-Rhin* and *Vosges*[16] changed. For this period, I distributed births of the old-classification *départements* between the unified-classification *départements*. The second change concerns the 1901–2014 period. It follows the *Ile-de-France* reorganization in 1964, effective in 1968. This reorganization led to the dissolution of *Seine* and *Seine-et-Oise*. These *départements* were divided between *Paris, Yvelines, Essonne, Hauts-de-Seine, Seine-Saint-Denis, Val-de-Marne* and *Val d'Oise.*

The missing data in the unified departmental classification are also of two types. The first concerns the missing data due to the two World wars: *Aisne, Ardennes, Marne, Meurthe-et-Moselle, Meuse, Nord, Oise, Pas-de-Calais, Somme* and *Vosges* for the 1914–1918 period, and *Moselle*, *Bas-Rhin* and *Haut-Rhin* for the period 1939–1945. *Corse* is also concerned in 1943 and 1944. The second category is *départements* temporarily under German control: this is the case of *Bas-Rhin*, *Haut-Rhin* and *Moselle* before 1919.

5.3.2.4 Specific Methods Due to Missing Data

Births of the missing *départements* during the period 1853–1900 are first estimated. Recall that these births allow the distribution of deaths according to Lexis triangles. I consider that the changes were synchronized between missing *départements* and a neighboring *département*. For *Var* and *Alpes-Maritimes*, whose limits are stable since 1861, I use the ratio between births in 1861 and births in *Bouches-du-Rhône* to deduce births between 1853 and 1860. I proceed in the same way for *Savoie* and *Haute-Savoie*, for which I use *Ain* as reference. Regarding *Vosges*, *Territoire de Belfort* and *Meurthe-et-Moselle*, I used *Haute-Saône* as reference for the 1853–1869 period. As I know values for *Meurthe, Moselle, Haut-Rhin* and *Vosges* (former *départements*), it was easy to deduce values for *Moselle* and *Haut-Rhin* in their current boundaries. For the 1870–1900 period, births in *Moselle*, *Bas-Rhin* and *Haut-Rhin* were estimated using *Haute-Saône* as reference.

Data from the population movement for missing *départements* during the two World wars are also estimated. Even if the lifetables of these *départements* should be analyzed with caution, this allows an approximation of their current mortality conditions. For that, I go further than the method used for births by endogenizing the choice of the reference *département*. For each couple of *département* and missing period, I choose a panel of geographically close *départements* whose data are available. Table 5.4 gives these candidates for each set of missing *départements*. I then calculate a score based

[14]Until 1870, two departments existed, namely *Meurthe* and *Moselle*. Their gathering fell within the same limits as *Meurthe-et-Moselle* and the new *Moselle*. The new *Moselle* includes the territories under German control in 1870, namely the districts of *Château-Salins* and *Sarrebourg* for the old *Meurthe* and T*hionville, Metz, Forbach-Boulay Moselle* and *Sarreguemines* for the old *Moselle*. In contrast, the new *Meurthe-et-Moselle* includes the territories remained French at that time, i.e. the districts of *Luneville*, *Nancy* and *Toul* for the old *Meurthe* and the canton of *Briey* for the old *Moselle.*

[15]In 1870, *Haut-Rhin* in its former boundaries is divided between *Haut-Rhin* as we know today – which passes under German control until the end of the Second World War – and *Territoire de Belfort*, which remains under French control.

[16]In 1870, the former cantons of *Schirmeck* and *Saales* (in *Vosges*) are attached to *Bas-Rhin*, which passes under German control. The new boundaries of these two *départements* are those that we know nowadays.

on the synchronicity of demographic variations over the period surrounding the missing period. From this score, a reference *département* is defined for each *département* with missing data and used to estimate these values. This method is used to both total births and deaths by age (sum of civilian, military and in deportation deaths). Section 5.6.6 of the Appendix goes into detail about the choice of reference *département* and the method used.

Table 5.4: Panel of candidate reference départements

Period	Missing *départements*	Panel of reference *départements*
1914–1919	*Aisne, Ardennes, Marne, Meurthe-et-Moselle, Meuse, Nord, Oise, Pas-de-Calais, Somme, Vosges*	*Aube, Eure, Haute-Marne, Haute-Saône, Seine-Inférieure, Seine-et-Marne, Seine-et-Oise*
1939–1945	*Moselle, Bas-Rhin, Haut-Rhin*	*Doubs, Meurthe-et-Moselle, Haute-Saône, Vosges*
1943–1944	*Corse*	*Alpes-Mar., Bouches-du-Rhône, Gard, Hérault, Var*

With the reorganization of *Ile-de-France* in 1968 I must differentiate the *départements* belonging to the old classification from those belonging to the new ones. The former *départements* are followed over the 1901–1968 period, and the new ones between 1968 and 2014. As such, I make several adjustments. The first concerns the distribution of births before 1968 among the *départements* of the new classification, in order to distribute deaths in Lexis triangles. It is done pro-rata 1968's births. Then I estimate 1968's age-populations for *départements* of the old classification by using the "Intercensal Survival" method: I assume that *Ile-de-France* migratory profile was the same for *Seine* and *Seine-et-Oise*. Section 5.6.7 of the Appendix discusses these two adjustments.

Finally, computation periods vary by *département*. I distinguish them according to four classes. Class 1 (C_1) concerns all *départements* outside *Moselle, Bas-Rhin, Haut-Rhin* and *Ile-de-France* (except *Seine-et-Marne*). These 85 *départements* are tracked over the period 1901–2014. Computations of population at each 1^{st} January is done as shown in Figure 5.3.2. *Départements* in class 2 (C_2) are the former *Ile-de-France départements*, namely *Seine* (75) and *Seine-et-Oise* (78). Lifetables were estimated over the period 1901–1968. Class 3 (C_3) concerns the new *Ile-de-France départements*: *Essonne* (91), *Hauts-de-Seine* (92), *Seine-Saint-Denis* (93), *Val-de-Marne* (94), *Val d'Oise* (95), *Paris* (96), *Yvelines* (97). Lifetables are available for the period 1968–2014. *Bas-Rhin, Haut-Rhin* and *Moselle* are in class 4 (C_4): lifetables are estimated between 1921 and 2014. Figures in Section 5.6.8 of the Appendix draw the methods used to estimate the January 1^{st} populations for each of these four classes. These are variants of Figure 5.3.2.

5.3.3 Reliability of the Data and Comparison with Other Studies

The raw data used in this study come from old statistical sources. I therefore verified that their use could be done without introducing bias in future analyzes.

Firstly, I was interested in the consistency of departmental and national data. Vallin and Meslé (2001) calculated the national lifetables for the 19th and 20th centuries. Consequently, I verified that the departmental sums of deaths, births, false stillbirths and populations are equal to national values. These expectations were true, which testify to the quality of the raw data. My results are therefore consistent with the results established at the national level.

Second, I was interested in the coherence of my results with the works already done at the departmental level. To do so, I calculated the differences between the departmental life expectancies of my paper and those of Bonneuil (1997) and Daguet (2006). Results are presented in Table 5.5.

Bonneuil (1997) calculated the life expectancies of women in 1901–1905. I have calculated life expectancies for the same period as well. The comparison between these estimates shows that mine are on average higher: the median of the difference is 3.34%. In addition, 50% of *départements* have a difference between 0.49% and 6.05%, and 25% of them have a difference of more than 6.05%. The in-depth study of age-specific mortality rates reveals that these differences are largely explained by lower infant mortality rates (deaths under age 5). Nevertheless, since I cannot retrieve the death and population statistics of Bonneuil (1997), I do not know if this difference comes from an underestimation of the number of deaths or an overestimation of the population at risk.

Daguet (2006) also revealed the departmental life expectancies at birth at the date of each census between 1954 and 1999. I compute the differences for both men and women. Overall, differences are much smaller. The median is around 0.2%, with no distinction for men and women and no temporal trend. The differences for 50% of the *départements* fall between 0% and 0.7% in 1962. These differences in 1999 for men are 0.22% and 0.73%, respectively. Although slight differences remain, one can conclude that life expectancies are reliable, even if slightly overestimated.

Table 5.5: Differences of departmental life expectancies at birth with other studies

	Men			Women		
	1^{st} Quart.	Med.	3^{rd} Quart.	1^{st} Quart.	Med.	3^{rd} Quart.
1901–1905				0.49	3.34	6.05
1954	0.18	0.65	1	0.54	0.84	1.34
1962	0	0.4	0.72	-0.01	0.37	0.68
1968	0.17	0.38	0.73	-0.02	0.33	0.78
1975	-0.17	0.15	0.5	-0.11	0.19	0.47
1982	0.01	0.27	0.59	0.04	0.21	0.5
1990	0.09	0.31	0.55	0.21	0.4	0.62
1999	0.22	0.49	0.73	0.47	0.66	0.99

Notes: Differences in % of my computations. Distribution of 90 or 95 departmental differences, according to the classification of the year.

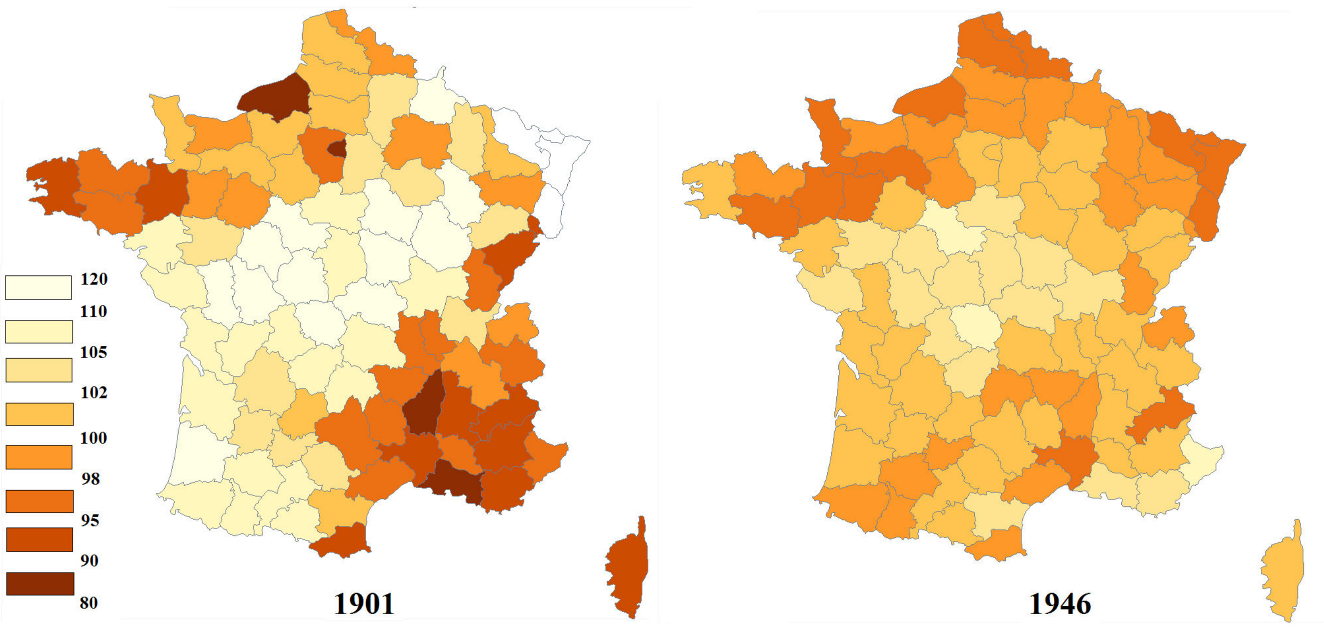
5.4 Available Results and Discussion

5.4.1 Available Results

Results are available for the 97 metropolitan *départements* monitored over the period 1901–2014, namely the *départements* of the current classification (*Corse* counting as one) as well as the old *Seine* and *Seine-et-Oise*. Due to their additivity, results are also available at the regional level in the classification prior to January 2016 (22 regions). The variables available are the life expectancies at each age (e_x) as well as a set of lifetable variables between ages 0 to 105 and over (number of survivors, mortality rates, proportions of deaths). Yearly births and populations by age are also available.

Figure 5.4.1 reveals the departmental life expectancies at birth relative to the metropolitan average, for women. I chose to present the results for women, but these results are available for men too. The first map shows the results for 1901. One can see that the highest life expectancies were located on an axis connecting the South-West to the North-East, from *Ardennes* to *Landes*. Maximums were reached in *Ardennes* but also in *Pays de la Loire* (*Loir-et-Cher, Indre, Indre-et-Loire, Deux-Sèvres, ...* etc.) and *Bourgogne* (*Côte d'Or, Yonne, Nièvre, ...* etc.) with values 10 to 20% higher than the metropolitan average. In contrast, life expectancies at birth in the South-East, *Seine* and *Bretagne* are significantly lower than the metropolitan average (between 5 and 20% according to the *département*). The second map presents these life expectancies at birth in the aftermath of the Second World War. At that time, maximums were reached in *Loir-et-Cher*, *Creuse* and *Alpes-Maritimes* with life expectancies 5 to 10% higher than the metropolitan average : Central-West was still a leader region, while the regions of *Bretagne* and *Normandie* were still lagging behind.

Figure 5.4.1: Relative life expectancy at birth, 1901 and 1946

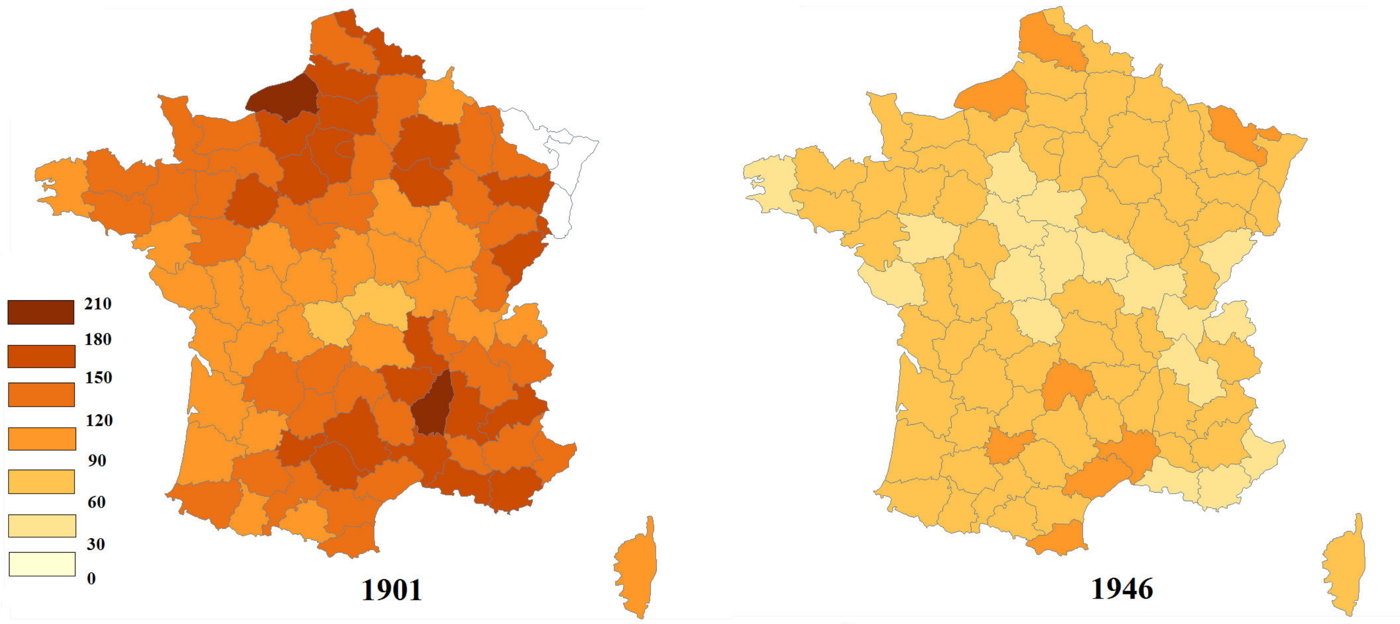


Notes: Life expectancy at birth, for women, in % of the metropolitan average. Sample includes 90 *départements*. *Moselle, Bas-Rhin and Haut-Rhin* values are non available in 1901 (*départements* under German administration).

Rather than analyzing synthetic indicators such as life expectancy, one can look at age-specific indicators. Since they impacted strongly life expectancies at birth, Figure 5.4.2 presents infant mortality rates for women. One more time I chose to present the results for women, but these results are available for men too. I represent the rates per thousand, and no longer relative to the metropolitan average. The landscape in 1901 was relatively similar to the map of life expectancy, since infant mortality rates were in 1901 very high. One can see that in extreme cases (*Seine-Inferieure, Ardèche*), for a thousand children under one year, between 180 and 210 died before their first birthday.

Rates were generally high in the North and the South-East (between 120 and 150), while they were lower in a broad central band connecting the *Saône-et-Loire* to the *Charente-Maritime* and the Atlantic coast. Minimums (between 60 and 90) were reached in *Creuse* and *Allier*. The second map shows the same values in 1946. Infant mortality rates decreased between the two years since they were globally around 60 per thousand in 1946. An under-mortality zone was visible, from *Eure-et-Loir* to *Isère* via *Nièvre*. The Mediterranean coast presented diverse situations: early mortality was low in the East (*Var, Alpes-Maritimes*) and strong in the West (*Hérault, Gard, Pyrénées-Orientales*).

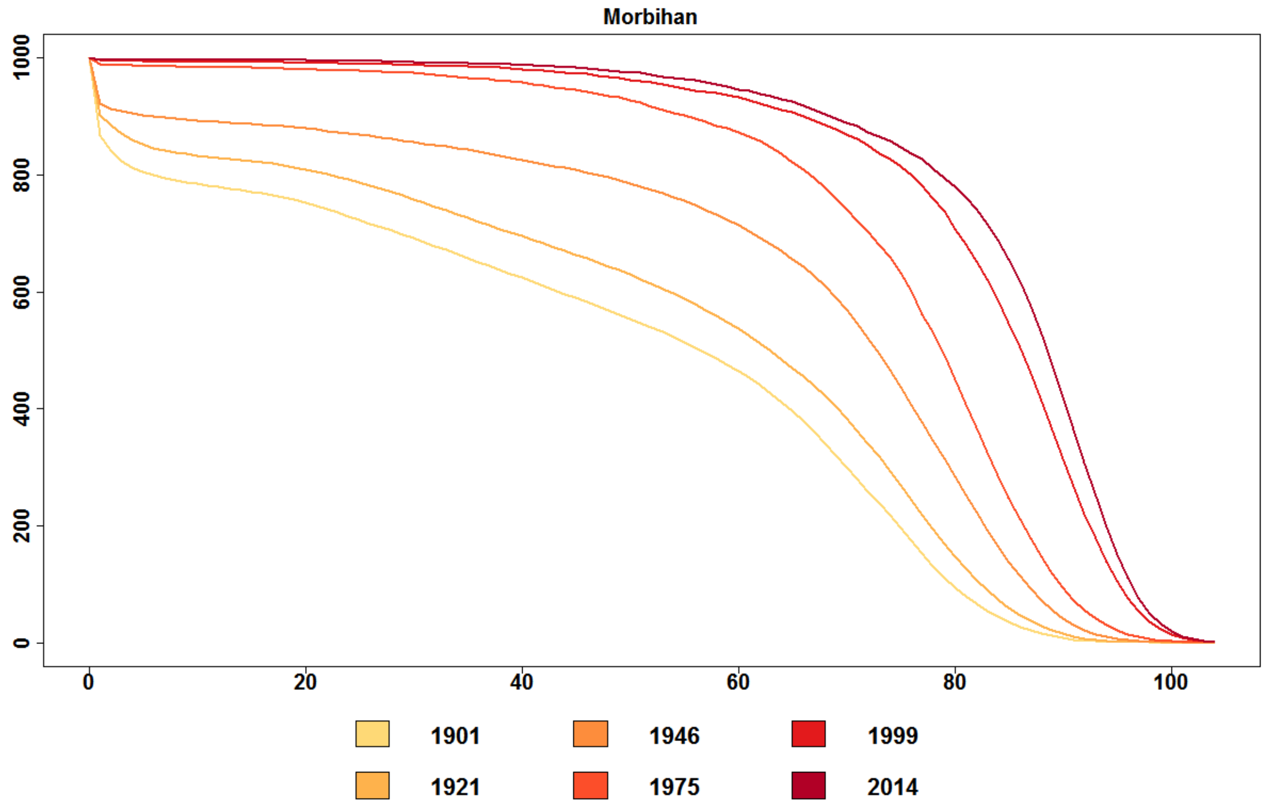
Figure 5.4.2: Infant mortality rates, 1901 and 1946



Notes: Infant mortality rates, for women, in thousand. Sample includes 90 *départements*. *Moselle, Bas-Rhin and Haut-Rhin* values are non available in 1901 (*départements* under German administration).

Finally, one can analyze evolvements of a single *département* over the 1901–2014 period. Figure 5.4.3 shows female survivors at each age for different dates in *Morbihan*. I have chosen this *département* since it was a place or high mortality in 1901. Indeed, there was high infant mortality at that time: there were only 850 survivors in the fictitious cohort. This infant mortality almost completely disappeared in 1975. The survival curve shifted to the upper-right corner as mortality rates were globally declining. This displacement was important until 1975, mainly because of the drop in infant mortality. Subsequently, the curve moved mainly because of the decrease in mortality between 60 and 80 years, then beyond 80 years for the 1999–2014 period. This is in line with the literature about rectangularization of the survival curve (see Wilmoth and Horiuchi (1999), Fries (2002), Cheung et al. (2005) for example): this curve was in 2014 very flat until age 60 (there is almost no death below this age). Beyond this age the curve decreases dramatically, especially beyond age 80.

Figure 5.4.3: Evolution of survivors at each age in morbihan



Notes: Survivors at each age, for women.

5.4.2 Discussion

5.4.2.1 Census Reliability

With population censuses one know the spatial distribution of the population by age and sex between the French *départements* along the 20th century. During this period, censuses served as a support for some public choices. The first concerns local budgets: allocations coming from central administration were based on the population of each territory. These censuses therefore affected the spatial distribution of public finance. The second concerns the electoral divisions: in order to obtain a fair representation in local or national assemblies, electoral divisions are divided so that each of them represents roughly the same population percentage. Censuses therefore had a very strong political impact. As a result, some regions have sought to inflate their census populations in order to get greater financial or electoral weight. Historians and statisticians have shown that *Marseille*'s population was overestimated in the 1930s.[17] This was also true in *Corse* in 1962: results of the exhaustive counting were not published because of inconsistencies. These censuses are, however, the basis of age-population computations. Even though ambiguous cases remain marginal over the period, they nevertheless existed.

[17]See *Statistique Annuelle du Mouvement de la Population,* 1939–1942, page 4

5.4.2.2 Interdepartmental Migrations

Methods used in this study partly take into account the issue of migrations. At each census date, the difference between estimated and recorded population can be seen as an approximation of net migration flows at each age. These flows are then distributed in proportion to the time elapsed between the first census and January 1^{st} of each year of the intercensal period. This approximation does not affect our results when the flows are weak or if they follow the approximation used. This is not the case in war periods. The May-June 1940 Exodus is an emblematic example. To escape the advance of German troops on French territory, the populations of the North-East migrate in mass towards the South and the West. I cannot take into account this exodus with the methodology used: the population of *Ardennes* on January 1^{st}, 1941 is for example largely overestimated. This issue is presented on several occasions in the *Statistique Annuelle du Mouvement de la Population* between 1939 and 1942[18]; this publication has suggested to estimate the present population with ration tickets dispensed to the population. However, Alary et al. (2006) showed that these tickets were circumvented during the war, questioning their reliability in counting the present population. Bonnet (2018) try to estimate these departmental populations, but only for females and for the total population.

Another issue relating to interdepartmental migrations concerns the nursery of children born in urban *départements*. Newborns were sent to rural *départements* close to major urban centers. Thus, *Seine* has a lower infant mortality rate than it should be because some of the infants are sent to suburbs. To overcome this issue, official publications suggests[19] to divide deaths of children under age 1 born in a *département* and living anywhere on the national territory, by the total of births in this *département*. I cannot do this because I do not find these raw data in official publications; this suggests that my infant mortality rates are slightly underestimated in urban *départements*.

5.4.2.3 Domiciliation of Deaths during the Two World Wars

The sources I use to estimate life expectancies during the two World Wars are incomplete: military and deportee deaths were recorded by birth *département* and not by home *département*. I build matrices linking birth *département* and home *département* before the deportation; nevertheless, they rely on strong assumptions about the representativity of pre- and post-war situations concerning the phenomena that took place during the war. The few statistics kept for this period limit the possibilities to go further. Regarding military deaths, I assume that the home *département* was similar to the birth *département* concerning the *"Morts pour la France"*. If this hypothesis seems weaker than those assumed for deportees, it is not entirely satisfactory. Again, I miss reliable and available data to overcome this issue.

5.4.2.4 Small *Département* Figures

Estimating fertility or mortality rates is difficult when figures are small (namely around 0). Papers tackle this issue by using bayesian estimation process (Asunção et al. (2005), Schmertmann et al.

[18]See *Statistique Annuelle du Mouvement de la Population,* 1939–1942, pages 3-4, 47 and 55

[19]See *Statistique Annuelle du Mouvement de la Population,* 1939–1942, pages 55

(2014) for fertility rates, Alexander et al. (2017) for mortality rates). The question arose of using these methods to supplement the HMD Protocol. However, the French *départements* figures are not as small as geographical units used in these studies: for example, the minimum according to population was reached in *Territoire de Belfort* in 1901 with 50,000 women, compared to 2,000 for some counties. However, these estimation models may be applied in the future, particularly to compute confidence intervals around departmental life expectancies.

5.5 Conclusion

In this paper, I have presented the sources and methods used to estimate lifetables by sex for all French metropolitan *départements* from 1901 to 2014. To do so, I have collected vital records and census statistics at the departmental level since the beginning of the 20th century. Since the two World Wars afflicted France between 1914–1918 and 1939–1945, military deaths and deaths in deportation were of great importance in the lifetables estimates; these statistics have been collected at the departmental level in original sources, namely the "*Mémoire des Hommes*" and "*MemorialGenWeb*" databases.

To estimate departmental lifetables, I have refered to the methods used in a large number of countries by the researchers of the Human Mortality Database. These methods transform the collected raw data into homogeneous data. They include the use of Cubics Splines to estimate deaths by age groups, the Kannisto model to extrapolate deaths at older ages, and a panel of methods to estimate populations at 1^{st} January of each year. The HMD protocol has been amended to take into account the French data specificities. This concerns false stillbirths which are reintroduced in the statistics of births and infant deaths, and territorial breaks such as those which affected the Paris region in 1968.

This work provides a new database on departmental mortality for the entire 20th century. Coupled with Bonneuil (1997)'s estimations for the 19th century, it provides an overview of the local trends in mortality since the French Revolution. As they have been calculated for each sex, these data shed new insights on the reasons explaining the differences in life expectancy between men and women . Moreover, beyond mortality statistics, this new database can be used to analyze all demographic fields at local level: birth rates since it includes annual births, the spatial distribution of population since it provides yearly populations by age, and finally internal migrations. These fields of research are on my future agenda.

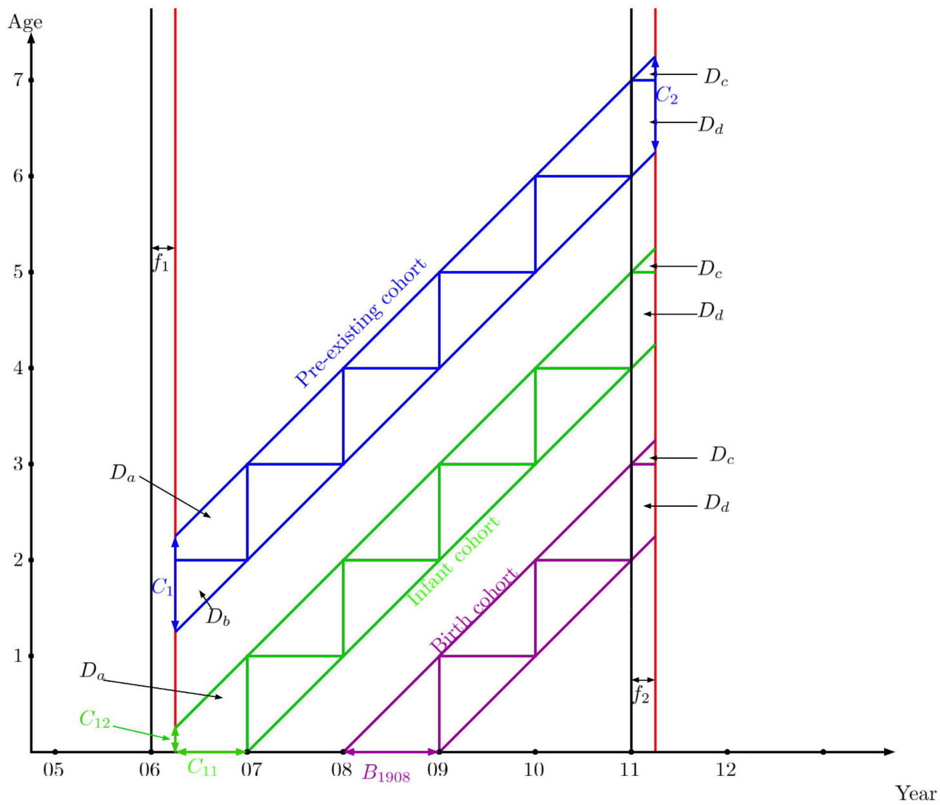
5.6 Appendices

5.6.1 Computations of Population on 1^{st} January

5.6.1.1 Intercensal Survival

The first method used to compute populations on 1^{st} January of each year is "Intercensal Survival". With this method I can estimate population by age for each intercensal period. Populations at the second census (e.g. 1906 for 1901–1906) are not estimated in the same way for all cohorts. Figure 5.6.1 presents the three types of cohorts which exist in this method. There are "Pre-existing cohorts" (born before the census year), "Infant cohort" (born during the census year) and "Birth cohorts" (born after the census year). The gaps between the census date and 1^{st} January of the census year are crucial. This gap is called f_1 for the first census and f_2 for the second.

Figure 5.6.1: Classification of different cohorts for Intercensal Survival method



I begin with "Pre-existing cohorts". I estimate age-population at date of the second census. Let t and $t+N$ be the first and last 1^{st} January in the intercensal period. N is the number of full calendar years between censuses. The dates of the two censuses are:

$$t_1 = t - 1 + f_1,$$

$$t_2 = t + N + f_2.$$

The elapsed time between the censuses is thus:

$$t_2 - t_1 = N + 1 - f_1 + f_2.$$

The cohort tracked (Figure 5.6.1, in blue) was 1- or 2-years-old at the time of the 1906 census and was born in 1904. Data are by year of birth and not by age, which simplifies computations. I assume a uniform distribution of deaths in each Lexis triangle, so that for the cohort aged x on 1^{st} January of the year of the first census,

$$D_a = (1 - f_1^2) \times D_L(x, t-1),$$

$$D_b = (1 - f_1)^2 \times D_U(x-1, t-1),$$

$$D_c = f_2^2 \times D_L(x+N+1, t+N),$$

$$D_d = (2f_2 - f_2^2) \times D_U(x+N, t+N).$$

This cohort's estimated population at the second census may be called $\hat{C}_2$ and is calculated as follows:

$$\hat{C}_2 = C_1 - (D_a + D_b) - \sum_{i=0}^{N-1} [D_U(x+i, t+i) + D_L(x+i+1, t+i)] - (D_c + D_d), \quad (5.6.1)$$

where $\Delta_x = C_2 - \hat{C}_2$ (the difference between the estimated population and that recorded at the date of the second census) comprises estimation errors and intercensal migrations within the cohort. In order to compute age-population at 1^{st} January of each intercensal year, the Δ_x error must be split between the age-populations in each intercensal year. I asume that these rough migrations are uniformly distributed over time. Population by age is calculated as follows:

$$P(x+n, t+n) = C_1 - (D_a + D_b) - \sum_{i=0}^{n-1} [D_U(x+i, t+i) + D_L(x+i+1, t+i)] + \frac{1 - f_1 + n}{N + 1 - f_1 + f_2} \Delta_x. \quad (5.6.2)$$

There is only one "Infant cohort" to track for each intercensal period (in Figure 5.6.1, the cohort born in 1906). Thus, $C_1 = C_{11} + C_{12}$, with $C_{11} = (1 - f_1) \times B_{t-1}$ and C_{12} the population recorded as born during the year of the census. Thus,

$$\hat{C}_2 = C_1 - D_a - \sum_{i=0}^{N-1} [D_U(i, t+i) + D_L(i+1, t+i)] - (D_c + D_d), \quad (5.6.3)$$

and

$$P(n,t+n) = C_1 - (D_a + D_b) - \sum_{i=0}^{n-1} [D_U(i,t+i) + D_L(i+1,t+i)] + \frac{\frac{1}{2}(1-f_1^2)+n}{N+\frac{1}{2}(1-f_1^2)+f_2}\Delta_0. \quad (5.6.4)$$

Finally, since N is the number of full calendar years during the intercensal interval, I track N birth cohorts. A cohort born in year $t+j$ is aged $K = N - j - 1$ on $01/01/t+N$. The estimated population of this cohort may be expressed as:

$$\hat{C}_2 = B_{t+j} - D_L(0,t+j) - \sum_{i=1}^{N-1} [D_U(i-1,t+j+i) + D_L(i,t+j+i)] - (D_c + D_d). \quad (5.6.5)$$

Note that the number of intermediate populations produced by the various cohorts depends on K. For $k = 0$, ..., K , the intermediate populations of each cohort are computed as follows:

$$P(k,t+j+k+1) = B_{t+j} - D_L(0,t+j) - \sum_{i=1}^{k} [D_U(i-1,t+j+i) + D_L(i,t+j+i)] + \frac{2k+1}{2K+1+2f_2}\Delta_{t+j}. \quad (5.6.6)$$

5.6.1.2 Precensal Survival Method

The second method I use is “Precensal Survival”, to compute populations for the first 1^{st} January of the whole period. Figure 5.6.2 presents the computations for population of age 1 in 1901. To do so, I must add $D_a^{'}$ et $D_b^{'}$ to the population born in 1901 and recorded on March 6^{th}, 1901. If t_1 is the first 1^{st} January of the intercensal period, then:

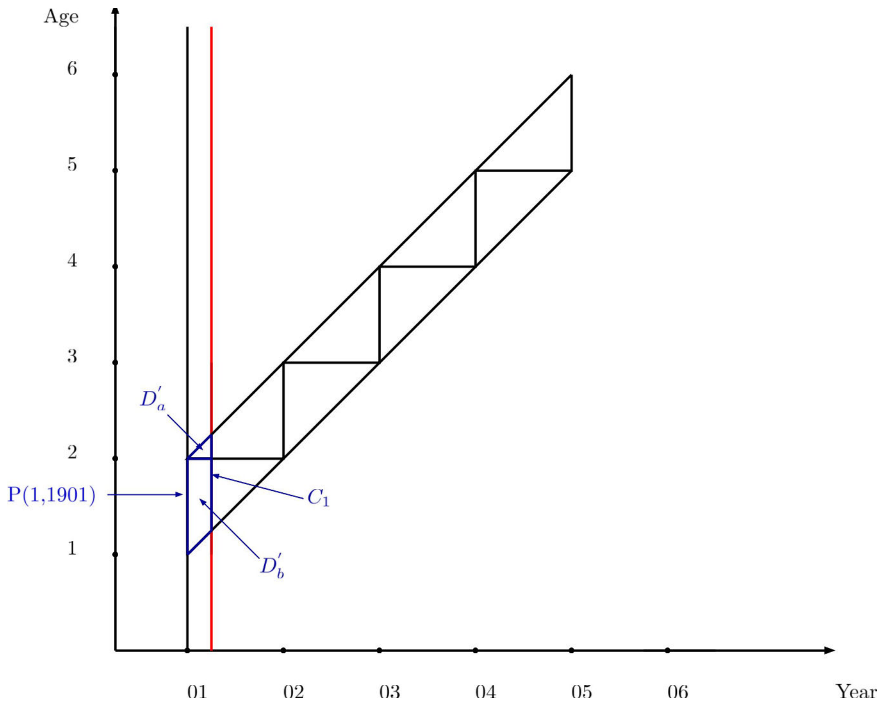
$$P(x-1,t_1-1) = C_1 + D_a^{'} + D_b^{'}. \quad (5.6.7)$$

5.6.1.3 Extinct Cohorts Method

The third method I use is “Extinct Cohorts”, to calculate age-population for the cohorts extincted in 2013. Since the maximum age in my database is 105, a cohort is considered to be extinct if it reached 105 or over in 2013. Figure 5.6.3 reveals that my data comprise two kinds of extinct cohorts. The first are “Full cohorts” (Figure 5.6.3, in red), which can be tracked from ages 80 to 105 in 1901–2013. Thus, the 80-year-old population in 1903 equals the sum of the cohort’s Lexis triangles between ages 80 and 105. The others are “Truncated cohorts” (Figure 5.6.3, in blue), those over age 80 in 1901. Thus, the 95-year-old population in 1901 equals the sum of the cohort’s Lexis triangles between 95 and 105. More generally, the population of age x in year t can be calculated as follows:

$$P(x,t) = \sum_{i=0}^{\infty} [D_U(x+i,t+i) + D_L(x+i,t+i)].$$

Figure 5.6.2: Precensal survival method



5.6.1.4 Survivor Ratio Method

The last method I use is "Survivor ratio", to calculate non-extinct cohorts of age 85 and over in 2013. Figure 5.6.4 presents the computations for the cohort aged 104 in 2013. The survivor ratio R may be defined as the number of individuals alive at age x on 1^{st} January t, divided by the number of individuals in the same cohort alive k years previously. Formally:

$$R = \frac{P(x,t)}{P(x-k,t-k)}.$$

I assume that there is no migration at these ages. R may also be expressed:

$$R = \frac{P(x,t)}{P(x,t)+\dot{D}}.$$

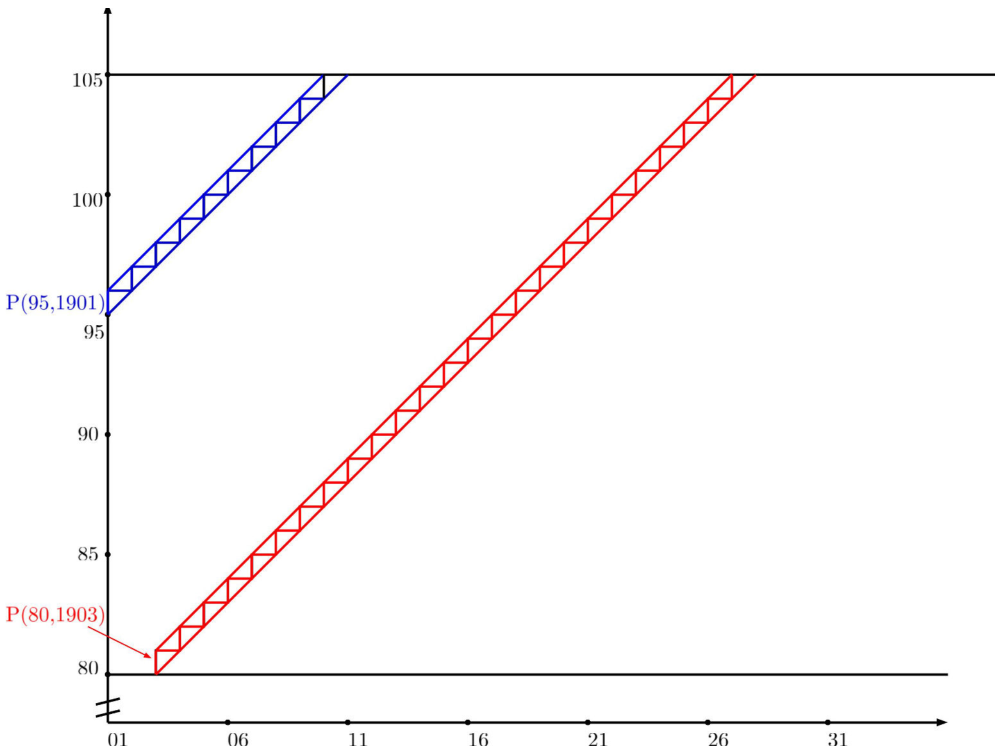
where$\dot{D} = \sum_{i=1}^{k}\left[D_U(x-i,t-i)+D_L(x-i+1,t-i)\right]$. Finally, $P(x,t)$ may be expressed as a function of R:

$$P(x,t) = \frac{R}{1-R}\dot{D}. \tag{5.6.8}$$

Since the survivor ratio cannot be directly observed for a cohort, I use preceding cohorts whose age-populations have been calculated by the "Extinct Cohorts" method. I asume that the survival ratio has roughly the same value in the studied cohort and in the preceding ones. As such, the mean ratio R^* of the preceding m cohorts may be calculated as follows:

$$R^*(x,2013,k,m) = \frac{\sum_{i=i}^{m} P(x,2013-i)}{\sum_{i=i}^{m} P(x-k,2013-k-i)}.$$

Figure 5.6.3: Extinct Cohorts Method



I may then estimate $\tilde{P}(x, 2013)$:

$$\widetilde{P}(x, 2013) = \frac{R^*}{1 - R^*}\dot{D}.$$

Subsequently, I may track the cohort back in time and estimate $\tilde{P}(x-1, 2012)$, $\tilde{P}(x-2, 2011), ...$ by adding step by step the cohort's deaths. I apply this method for any non-extinct cohort in 2013. For my estimations I follow the guidelines of the HMD Protocol, with $k = m = 5$.

The assumption of a constant survivor ratio over time is strong, and I may control by the recorded population on 1^{st} January 2013. I compare the 85-and-over population on 1^{st} January 2013 – retrieved from the census of that year (called P_{85+}^{Rec}) – with the 85-and-over population on 1^{st} January 2013 as calculated by the Survivor Ratio method (called P_{85+}^{SR}). Thus, populations at each age in 2013 can be computed as follows:

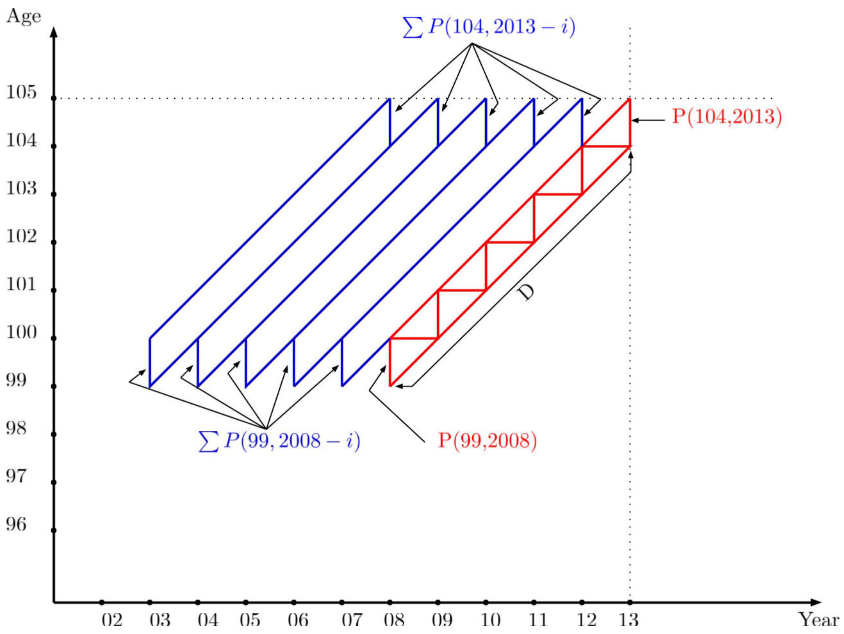
$$\hat{P}(x, 2013) = c\widetilde{P}(x, 2013) = c\frac{R^*}{1 - R^*}\dot{D},$$

where $c = \frac{P_{85+}^{Rec}}{P_{85+}^{SR}}$. As before, each cohort is back-followed: I make estimates for $\hat{P}(x-1, 2012)$, $\hat{P}(x-2, 2011), ...$

5.6.2 Set of Different Lifetables

Concerning computations of (1×1) and (1×5) lifetables, I start from values of q_x. With these values I compute p_x the probability of staying alive between x and $x + 1$. Then I compute the number of survivors at each age per 100,000 births

Figure 5.6.4: Survivor Ratio Method



$$l_x = l_0 \prod_{i=0}^{x-1} p_i \text{ with } (l_0 = 100,000),$$

the deaths at each age(d_x)

$$\begin{cases} d_x = l_x q_x & x \in (0, 104) \\ d_x = l_x & x = 105 \end{cases},$$

the number of years lived between x and $x + 1$

$$\begin{cases} L_x = l_x - (1 - a_x)d_x & x \in (0, 104) \\ L_{105}^{\infty} = l_x a_x & x = 105 \end{cases},$$

the number of life years remaining to live

$$\begin{cases} T_x = \sum_{i=x}^{104} L_x + L_{105}^{\infty} & x \in (0, 104) \\ T_x = L_{105}^{\infty} & x = 105 \end{cases}.$$

Finally, life expectancy at age x is computed as follows:

$$e_x = \frac{T_x}{l_x}.$$

Methods are quite the same for (1×5) lifetables. I therefore get lifetables for quinquennial periods: 1901–1905, 1906–1910, 1911–1915... etc. Values in abridged (5×1) and (5×5) lifetables are computed with previous variables. ${}_5e_x$, ${}_5l_x$ and ${}_5T_x$ are directly retrieved from the complete lifetables. Finally, ${}_5d_x = l_x - l_{x+5}$, ${}_5q_x = \frac{d_x}{l_x}$ and ${}_5L_x = T_x - T_{x+5}$. One can also find ${}_5a_x$ and ${}_5m_x$ from the basic formula linking all these variables.

5.6.3 Census Adjustments

For my purposes it is simpler to compute population figures by birth year. Census data are given by single age after 1968. I gather populations by five-year age groups between ages 15 and 89, before taking the open-age interval 90 and over. The cubic splines adjustment takes into account that populations were given by age and not by birth year. Thus, taking the 1968 census as an example, I isolated the populations born between 01/01/1968 and the date of the census.[20] Before 1968, data are given by birth year. Nevertheless some specific adjustments are needed.

5.6.3.1 Distribution of Deaths of Unknown Age in 1901 Census

For the 1901 census, individuals whose birth year is unknown are put together in the open-age interval. To allocate them I use the 1911 census, which has a useful degree of detail. The process follows three steps. The first is based on the calculation of the quotient of individuals aged 95 and over by individuals aged 80 and over for each *département* i and each sex j in 1911:

$$R_{95ij}^{1911} = \frac{\sum_{s=95}^{105} P_{sij}^{1911}}{\sum_{s=80}^{105} P_{sij}^{1911}}. \tag{5.6.9}$$

These quotients are then applied to the 1901 census to compute the proportion of individuals aged 95 and over among individuals aged 80 and over:

$$\sum_{s=95}^{105} P_{sij}^{1901} = R_{95ij}^{1911} \times \sum_{s=80}^{105} P_{sij}^{1901}. \tag{5.6.10}$$

By substraction, I finally deduce death of unknown year of birth for each *département* and sex.

5.6.3.2 Addition of Age Group for Pre-1946 Censuses

The 1906, 1921, 1926, 1931, 1936 and 1946 censuses did not use the same methodology for populations in the first three age groups. Some groups have to be combined or splitted (Table 5.6, in italics). For that purpose I assume that births were spread uniformly over time.

Finally, the 1911 census is rather different because it provides data for each year of birth and not per five-year groups. However, these numbers fluctuate considerably. There were two possible methods: either use the numbers given, or combine the numbers in five-year groups as for the other censuses and apply cubic splines. Although the first method provides more information, it includes inconsistent fluctuations at adult ages. Since I need to maintain consistency, I choose the second method. Raw data in 1911 have to be thoroughly reprocessed: I keep the first fifteen birth year groups, and then combine them by five-year groups (1891–1895, 1886–1890, etc.) plus the open-age interval "1820 and earlier".

[20] Note that the estimates of the population born in the census year are important because they are used to calculate populations by 1^{st} January of each year.

Table 5.6: CLASSIFICATION AND AVAILABILITY OF POPULATIONS BORN TWO YEARS BEFORE THE CENSUS

Census	1^{st} class	2^{nd} class	3^{rd} class
1901	Born from 01/01/01 to 04/03/01	Born in 1900	Born in 1899
1906	Born from 01/01/06 to 03/06/06	Born in 1905	Born in 1904
1911	Born from 01/01/11 to 03/05/11	Born in 1910	Born in 1909
1921	Born from 01/01/21 to 03/05/21	*Born from 03/06/20 to 12/31/20*	*Born from 01/01/20 to 03/05/20*
1926	Born from 01/01/26 to 03/07/26	*Born from 03/08/25 to 12/31/25*	*Born from 01/01/25 to 03/07/25*
1931	Born from 01/01/31 to 03/07/31	*Born from 03/08/30 to 12/31/30*	*Born from 01/01/30 to 03/07/30*
1936	Born from 01/01/36 to 03/07/36	*Born from 08/03/35 to 31/12/35*	*Born from 01/01/35 to 7/03/35*
1946	*Born from 03/10/45 to 03/09/46*	*Born from 01/01/44 to 03/09/45*	Born in 1943

Notes: Periods in italics in the table have to be combined or splitted to get populations by year of birth. 01/01/01 means 01/01/1901.

5.6.3.3 Adjustment of Censuses by Cubic Splines

To get populations by single year of birth and not five-year groups, I adjust census populations by cubic splines, as I do for civilian and military deaths. The cubic splines are fitted to the cumulative curve of population born before 1^{st} January of the census year. For example, according to the 1901 census, I consider the population born before 1^{st} January 1901. The population born between 1^{st} January 1901 and the day of the census provide no further information and would involve fractional knots.

5.6.4 Estimates of Military Deaths during the Two World Wars

The classification of *départements* from the *"Mémoire des Hommes"* website is modified to fit the classification for civilian deaths. Problems concern *Corse* (two *départements* counting as one) and the old *départements* of *Seine* and *Seine-et-Oise*. For these last two, deaths are given for the new *départements*. To allocate deaths between *Seine* and *Seine-et-Oise* I first sum all deaths in *Ile-de-France* (without *Seine-et-Marne*), then I allocate these military deaths pro rata of population in the cohorts born from 1880 to 1896. These cohorts account for 83% of total military deaths in the First World War. Concerning the distribution of deaths in the Parisian *départements* between *Seine* and *Seine-et-Oise* for the Second World War, I allocate them pro rata of populations born between 1905 and 1921 (70% of total deaths during the Second World War). *Seine*'s deaths are equal to 78.6% of the total.

Moreover, to ease the collection of data from the website, military deaths have been retrieved by year of birth for the youngest (born after 1889), then by five-year group for those born in 1889 and earlier. These deaths must be split by year of birth, which is done by cubic splines. The two assumptions made are (1) no deaths under age 16 and (2) no deaths over age 60.

5.6.5 Estimates of Deportees

The deportee database is nominative (1 line for each deportee). Sex, birth *département* (or country of birth if born abroad), day-month-year of birth, day-month-year of death were extracted. The age of death in days-months-years follows. For dates of birth and death, data are kept since the year was available. Thus, if only the year was available, the date chosen was January 1st. Likewise, if only the month and year of birth were available, the full date of birth was set to the first day of the month. If the date was considered irrelevant (namely if date of birth after date of death), the date is erased. For individuals whose year of death was after 1946 (for about forty individuals), I consider that those are unknown. 93% of the deceased have well-informed data for the four variables (sex, date of death, age, place of birth). For those with two or three variables missing, data were not used. This corresponds to 6.5% of the database. I did not use deportees with one variable missing too since they represented only 0.5% of the total. From these nominative data, I thus extract matrices crossing the age of death, the year of death (1940–1946), the place of birth and the sex.

One of the variables available in the deportee database is the place of birth. One has to differentiate this variable from the home place before deportation, that is where the deceased would have to be located in my lifetables. Since a 40-year-old have a non-zero probability to migrate in a different *département* from where he is born, I may infer the home-*département* before deportation. Similarly, deportees born abroad must be located in a French *département.*

5.6.5.1 Born-abroad Deportees

There are 33,609 deaths of born-abroad deportees, some 44% of the database. Those born outside France need to be allocated across France on the assumption that they immigrated before they were arrested and deported. One may suppose that these deportees born outside France fled Nazi persecution and settled in France before the start of the war. I make the assumption that the probability of being in each *département* can be infered by the spatial distribution of foreigners in 1936. Moreover, I assume that this distribution does not vary by age, and also that the 1936 distribution is representative of the war-time one. I can construct the following matrices:

1. N: $PaysN \times Age$ (90 × 105) (taken from the MemGenWeb base),
2. P: $DeptR \times PaysN$ (91 × 48) (taken from the 1936 census),
3. R: $DeptR \times Age$ (90 × 105).

The first modification concerns *Seine*. Matrix P comprises 91 *départements* and not 90 because of the distinction we make between the city of Paris and the inner suburbs, so these two lines are summed to get the same administrative boundaries in the two matrices. Next P must be transformed so that the matrix gives us the probability that an individual born in country i lives in *département* j. Each element in the matrix equals:

$$P^*_{ij} = \frac{T_{ij}}{\sum_i T_{ij}} = \frac{T_{ij}}{T_{.j}},$$

with T the total of deportees.

Third, the names of countries of birth for Matrices P and N must be linked: there are 48 countries or regions in Matrix P and 90 countries in Matrix N.[21] I need to reclassify them to calculate the product of Matrices N and P. Thus I get a Matrix $P^*(90 \times 90)$ and calculate the R Matrix:

$$R' = N'P^*$$

Ultimately, each element R_{sj} in Matrix R corresponds to the sum of individuals aged s born in each of the countries i who emigrated to département j before being arrested and deported.

5.6.5.2 French Deportees

There are 43,055 deaths of French-born deportees in the database. I cannot assume that any deportee born in a *département* stayed in that *département*. A transfer matrix must therefore be constructed linking *département* of birth and *département* of residence before deportation. I use the matrix cross-referencing *département* of residence and *département* of birth in the 1946 census.[22] I assume both this matrix is representative of the pre-war situation and of deportee migrations, and that the probability of migration is equal for all ages.

I make a few preliminary modifications. The main is to allocate the deportees according to the post-1968 *départements* between *Seine* and *Seine-et-Oise*. The allocation key is the same as the one used for military deaths in the Second World War. I construct the following matrices:

1. N: $DeptN \times Age$ (90 × 105),
2. P: $DeptN \times DeptR$ (90 × 90),
3. R: $DeptR \times Age$ (90 × 105).

P is transformed so that the matrix gives the probability that an individual born in *département* i lives in *département* j. Each element in the matrix equals:

$$P^*_{ij} = \frac{T_{ij}}{\sum_i T_{ij}} = \frac{T_{ij}}{T_{.j}}.$$

Thus I deduct:

$$R' = N'P^*.$$

The matrices of French and foreign-born deportees are finally added. This final matrix is the sum for each *département*, each age, each sex and each year, of the departees born in a French *département* and deportees born outside France but living in France when they were arrested. For the total number of deportees, I based my computations on Dupaquier (1988). He followed Sauvy, who reports 27,000

[21] The level of detail in the MemGenWeb database is quite high, whereas the one in the census is lower (many Asian, South American and African countries are not directly specified, and colonies are often included in the generic term "French possessions in Africa").

[22] This matrix distinguishes males and females, enabling us to refine the estimates.

resisters who died in deportation and 83,000 jewish and other deportees. Consequently, I consider that 110,000 individuals died in the camps.

5.6.6 Missing Data During the Two World Wars

There are ten *départements* (*Aisne, Ardennes, Marne, Meurthe-et-Moselle, Meuse, Nord, Oise, Pas-de-Calais, Somme, Vosges*) with missing data during the First World War, and four during the Second World War (*Corse* between 1943 and 1944, *Moselle, Bas-Rhin, Haut-Rhin* between 1939 and 1945). These missing data are of two types: births and stillbirths, as well as deaths. The general assumption used for the estimations of these missing data is that the neighbours *départements* have similar evolutions concerning their demographic variables, because of their culture and their shared living conditions. As a result, I can estimate the evolution of the missing demographic variables using a reference *département*.

5.6.6.1 Births and stillbirths

The choice of the reference *département* for each of the missing *départements* and each sub-period must consider how their demographic variables were synchronized. For that purpose, I may define a support interval and then track changes in the ratio between the variable in the missing *département* and in the reference *département* during that interval. Let t_1 and t_2 be the first and last years of the subperiod for which there are missing data, $\Omega_\Delta = [t_1, t_2]$ the subperiod for which there are missing data, $\varOmega_t = [t_1 - h, t_1] \cup [t_2, t_2 + h]$ the support interval with $h = 4$, i the missing *département*, j the potential reference *département*. The ratio R_{ij}^t is calculated for a demographic variable V:

$$R_{ij}^t = \frac{V_j^t}{V_i^t},\ t \in \varOmega_t \quad .$$

Then mean ($\bar{x}_{ij}$) and standard deviation (σ_{ij}) of R_{ij}^t are calculated over the interval $\varOmega_t$. The stability of the ratio is measured as the coefficient of variation of R_{ij}^t over the interval $\varOmega_t$:

$$CV_{ij} = \frac{\sigma_{ij}}{\bar{x}_{ij}}.$$

The reference *département* j^* chosen is the one with the lowest coefficient of variation among all the possible reference *départements*. This criterion is used for both stillbirths and births. Since births moves in a similar way for both sexes, the reference *département* is chosen by examining female births. The same choice is then applied to males and females. After choosing the reference *département* for each missing *département*, the missing data for *département* i and variable V is estimated as follows:

$$V_i^t = V_{j^*}^t \times \bar{x}_{ij*}.$$

5.6.6.2 Deaths

The method used to estimate missing deaths is similar to the one used for stillbirths and births. Note that computations are made for total deaths (including military deaths and deportees). Let t_1

and t_2 be the first and last years of the subperiod for which there are missing data, $\Omega_\Delta = [t_1, t_2]$ the subperiod for which there are missing data, $\Omega_t = [t_1 - h, t_1] \cup [t_2, t_2 + h]$ the support interval with $h = 4$, i the missing *département*, j the potential reference *département*. The ratio $R^t_{xij}$is calculated for deaths D at age x:

$$R^t_{xij} = \frac{D^t_{xj}}{D^t_{xi}} \; (t \in \Omega_t).$$

Then mean ($\bar{x}_{xij}$) and standard deviation (σ_{xij}) of R^t_{xij} are calculated over the interval Ω_t. The stability of the ratio is measured as the coefficient of variation of R^t_{xij} over the intervalΩ_t:

$$CV_{xij} = \frac{\sigma_{xij}}{\bar{x}_{xij}}.$$

The fit between missing *département* and reference *département* needs to take the lowest value of the coefficient of variation over a number of ages Ω_x and not a single point. I calculate a score S_{ij}:

$$S_{ij} = \frac{1}{\Omega_x} \sum_{x \in \Omega_x} CV_{xij},$$

where Ω_x is defined as ages 0–4 and 50–89 in order to avoid erratic results due to small number of deaths.

The reference *département* j^* chosen is the one with the lowest score among all the possible reference *départements*. After choosing the reference *département* for each missing *département* and subperiod, deaths at age x for the *département* i are estimated as follows:

$$D^t_{xi} = D^t_{xj^*} \times \bar{x}_{xij^*}.$$

5.6.7 Reorganization of *Ile-de-France* in 1968

By changing the three *départements* of *Ile-de-France* (*Seine, Seine-et-Marne, Seine-et-Oise*) in eight new ones (*Paris, Seine-et-Marne, Yvelines, Essonne, Hauts-de-Seine, Seine-Saint-Denis, Val-de-Marne, Val-d'Oise*), the reorganization of this region in 1968 creates a discontinuity in data. I change my methodology so as to track each of these *départements* over the most appropriate period. Note that *Seine-et-Marne* was not affected by these changes. When I talk about *Ile-de-France* hereafter, I mean the *Ile-de-France* region less *Seine-et-Marne*.

For the intercensal period 1901–1962, I can track the old *départements*: I have all the censuses between these years and population flows (births and deaths). For the intercensal period 1968–2014, I can track the new *départements*: I have all the censuses and population flows between those dates. For the intercensal period 1962–1968, I have 1962 and 1968 censuses for the new *départements*, but no population flows. For the same intercensal period, I have population flows and the 1962 census for the old *départements*, but no data according to the 1968 census. I choose to track the old *départements* until 1968, and the new ones from 1968 onwards. To do so, I make two adjustments. The first is about pre-1968 births for the new *départements*, useful to split deaths in Lexis triangles. The second is about populations of the old *départements* in 1968, to estimate the 1^{st} January population of these *départements* between 1962 and 1968.

To estimate births of the new *départements* before 1968, I use the 1968 distribution. I assume that the weight of each *département* remains constant. Although this is a strong assumption if one want to know the accurate number of births, it is less strong for the relative size of two successive cohorts.

I am not able to calculate 1^{st} January populations of the 1962–1968 intercensal period for *Seine* and *Seine-et-Oise.* Indeed, the turning census available for both old and new *départements* is the 1962 one. In order to estimate pre-1968 population, one need population aged 85 and over to implement the "Survivor Ratio" method, and populations aged 0 to 84 to implement the "Intercensal Survival" one. To estimate the population aged 85-and-over for *Seine* and *Seine-et-Oise*, I assume that the weight of the two *départements* in the Ile-de-France 85-and-over population did not vary between 1962 and 1968.

It is more difficult concerning the population aged 0 to 84. To do so, I draw on the Intercensal Survival method. First, I calculate the estimated population in 1968 for *Seine* and *Seine-et-Oise* and the sum of these two *départements* $\left(\hat{P}^{68}_{IdF}(x)\right)$, by subtracting from each cohort counted in 1962 deaths occurring during the intercensal period. I also know the population estimated for these two *départements* in 1968 ($P^{68}_{IdF}(x)$) by summing the new *départements.* I can therefore deduce the migratory profile for *Ile-de-France*:

$$R^{68}_{IdF}(x) = \frac{\hat{P}^{68}_{IdF}(x)}{P^{68}_{IdF}(x)}.$$

I assume this profile was similar for each of the old *départements* j and use this migratory profile to compute 1968 census populations:

$$P^{68}_{j}(x) = R^{68}_{IdF}(x) \times \hat{P}^{68}_{j}(x).$$

5.6.8 Computations of 1^{st} January Populations by Class of *Département*s

I miss data for some *départements* between 1901 and 2014 in order to compute lifetables. Consequently, I divided my panel into four classes:

C_1 All départements except *Moselle, Bas-Rhin, Haut-Rhin, Seine-et-Oise* and *Ile-de-France* (except *Seine-et-Marne*). These *départements* are tracked between 1901 and 2014. Figure 5.3.2 presents how I calculate populations at each 1^{st} January.

C_2 *Seine* (75) and *Seine-et-Oise* (78). The lifetables for these *départements* are estimated for the period 1901–1968. Figure 5.6.5 presents the methods used to compute populations at each 1^{st} January. One can see that the Survivor Ratio method is applied to the 1968 census and not the 2013 census.

C_3 The new *départements* in *Ile-de-France: Essonne* (91), *Hauts-de-Seine* (92), *Seine-Saint-Denis* (93), *Val-de-Marne* (94), *Val d'Oise* (95), *Paris* (96), Yvelines (97). These lifetables are estimated between 1968 and 2014. Figure 5.6.6 presents the methods used to compute populations at each 1^{st} January.

Figure 5.6.5: Estimations of populations for départements of class 2

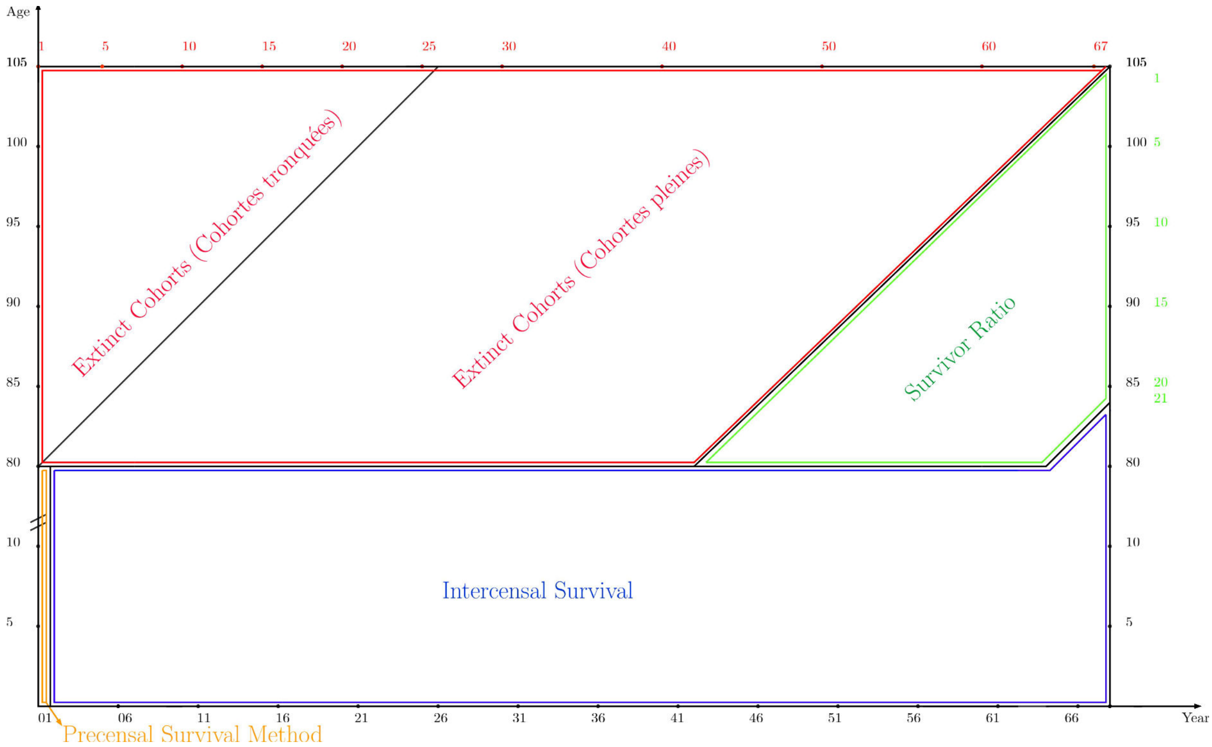
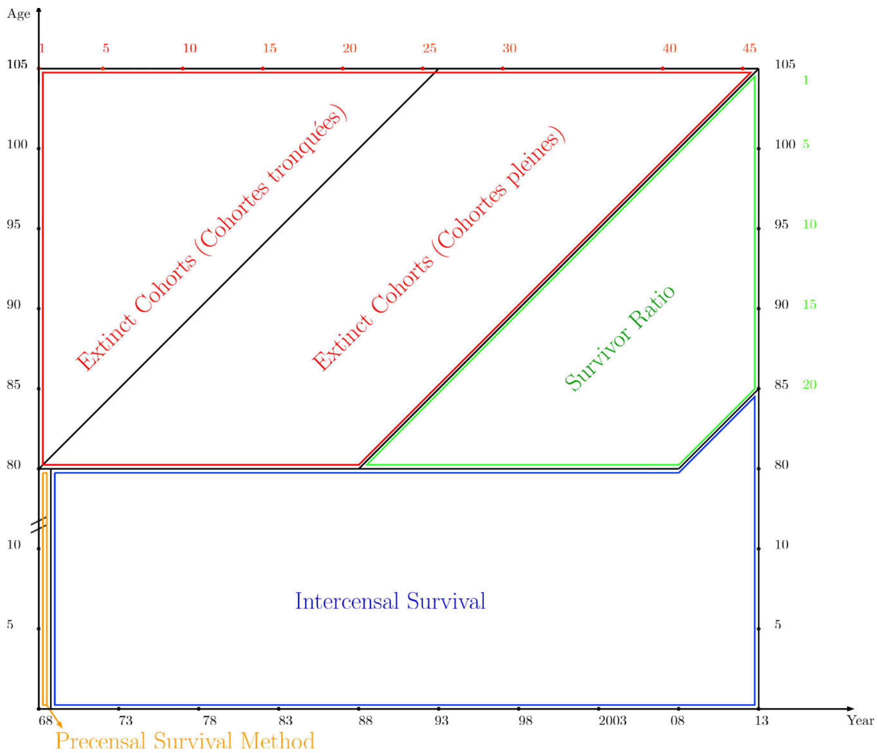
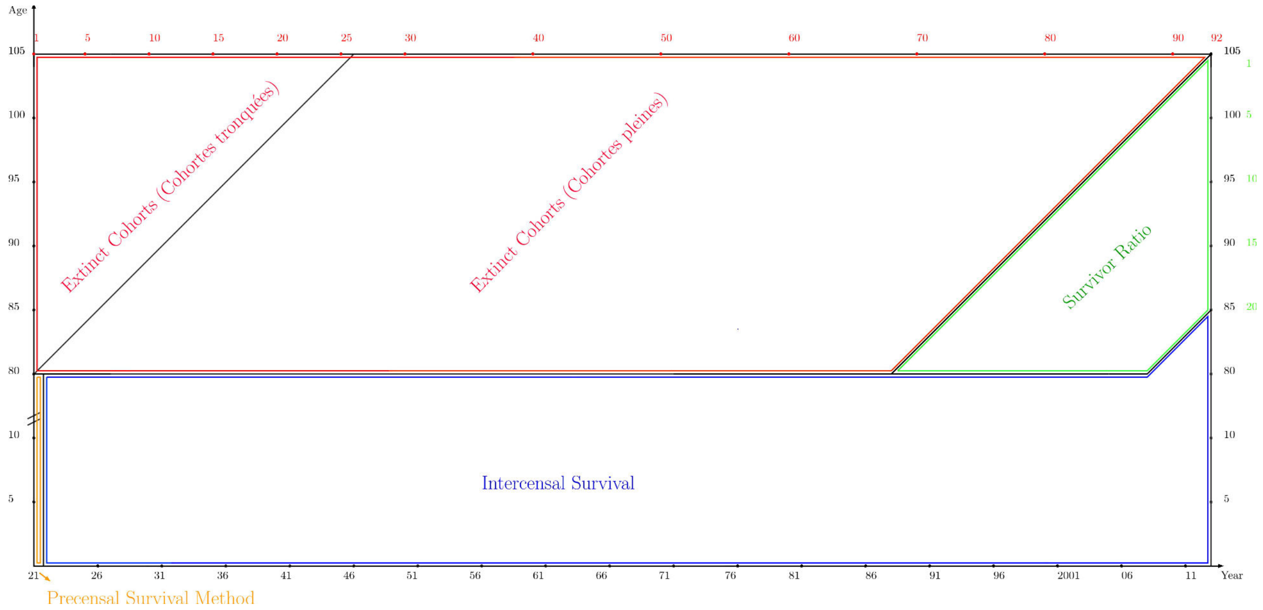


Figure 5.6.6: Estimations of populations for départements of class 3



C_4 *Moselle, Bas-Rhin* and *Haut-Rhin*. From 1870 to 1918 these three *départements* were under German administration. Consequently, the public records were not kept by the French authorities. I have not been able to do research in Germany to find data for this territory, so my estimates begin at the first available census, namely 1921, as shown in Figure 5.6.7.

Figure 5.6.7: Computations of populations for départements of class 4



5.6.9 Sources of Raw Data

Table 5.7: Sources for civilian deaths, 1901–1929

Year	Départements		Classes		Publication	Book	Page
	Total	Missing	Step	Ceiling			
1901	87	(2)	5	100+	SAMP (Year 1901)	31	62–73
1902	87	(2)	5	100+	SAMP (Year 1902)	32	62–73
1903	87	(2)	5	100+	SAMP (Year 1903)	33	72–83
1904	87	(2)	5	100+	SAMP (Year 1904)	34	62–73
1905	87	(2)	5	100+	SAMP (Years 1905 et 1906)	35-36	62–73
1906	87	(2)	5	100+	SAMP (Years 1905 et 1906)	35-36	140–144
1907	87	(2)	5	100+	SAMP (Years 1907-1908-1909-1910)	1	190–193
1908	87	(2)	5	100+	SAMP (Years 1907-1908-1909-1910)	1	194–197
1909	87	(2)	5	100+	SAMP (Years 1907-1908-1909-1910)	1	198–201
1910	87	(2)	5	100+	SAMP (Years 1907-1908-1909-1910)	1	202–205
1911	87	(2)	5	100+	SAMP (Years 1911-1912-1913)	2	152–155
1912	87	(2)	5	100+	SAMP (Years 1911-1912-1913)	2	156–159
1913	87	(2)	5	100+	SAMP (Years 1911-1912-1913)	2	160–163
1914	77	(1)	5	100+	SAMP (Years 1914–1919)	3	100–103
1915	77	(1)	5	100+	SAMP (Years 1914–1919)	3	104–107
1916	77	(1)	5	100+	SAMP (Years 1914–1919)	3	108–111
1917	77	(1)	5	100+	SAMP (Years 1914–1919)	3	112–115
1918	77	(1)	5	100+	SAMP (Years 1914–1919)	3	116–119
1919	77	(1)	5	100+	SAMP (Years 1914–1919)	3	120–123
1920	90		5	100+	SAMP (Years 1920–1924)	4	82–85
1921	90		5	100+	SAMP (Years 1920–1924)	4	86–89
1922	90		5	100+	SAMP (Years 1920–1924)	4	90–93
1923	90		5	100+	SAMP (Years 1920–1924)	4	94–97
1924	90		5	100+	SAMP (Years 1920–1924)	4	98–101
1925	85		5	100+	SAMP (Year 1925) - CD	5	2–183
1926	90		5	100+	SAMP (Year 1926) - CD	6	2–183
1927	90		5	100+	SAMP (Year 1927) - CD	7	2–183
1928	90		5	100+	SAMP (Year 1928) - CD	8	2–183
1929	90		5	100+	SAMP (Year 1929) - CD	9	2–183

Notes: SAMP: *Statistique Annuelle du Mouvement de la Population*; CD: *Causes de Décès*.

(1) *Aisne - Ardennes - Marne - Meurthe et Moselle - Meuse - Moselle - Nord-Oise - Pas de Calais - Bas Rhin - Haut Rhin - Somme - Vosges*
(2) *Moselle - Bas Rhin - Haut Rhin*

Table 5.8: Sources for civilian deaths, 1930–2014

Year	Départements		Classes		Publication	Book	Page
	Total	Missing	Step	Ceiling			
1930	90		5	80+	SAMP (Year 1930) - CD	10	16–195
1931	90		5	80+	SAMP (Year 1931) - CD	11	16–195
1932	90		5	80+	SAMP (Year 1932) - CD	12	16–195
1933	90		5	80+	SAMP (Year 1933) - CD	13	16–195
1934	90		5	80+	SAMP (Year 1934) - CD	14	16–195
1935	90		5	80+	SAMP (Year 1935) - CD	15	16–195
1936	90		5	80+	SAMP (Year 1936) - CD	16	16–195
1937	90		5	100+	SAMP (Year 1937)	17	54–57
1938	90		5	100+	SAMP (Year 1938)	18	154–157
1939	87	(2)	5	100+	SAMP (Years 1939–1942)	19	118–125
1940	87	(2)	5	100+	SAMP (Years 1939–1942)	19	178–185
1941	87	(2)	5	100+	SAMP (Years 1939–1942)	19	238–245
1942	87	(2)	5	100+	SAMP (Years 1939–1942)	19	298–245
1943	86	Corse + (2)	5	100+	SAMP (Year 1943)	20	58–65
1944	86	Corse + (2)	5	100+	SAMP (Year 1944)	21	58–65
1945	87	(2)	5	100+	SAMP (Year 1945)	22	60–67
1946	90		5	100+	SAMP (Years 1946–1947)	23	110–117
1947	90		5	100+	SAMP (Years 1946–1947)	23	170–177
1948	90		5	100+	SAMP (Years 1948–1949)	24	242–249
1949	90		5	100+	SAMP (Years 1948–1949)	24	308–315
1950	90		5	100+	SAMP (Years 1950–1951)	25	240–247
1951	90		5	100+	SAMP (Years 1950–1951)	25	314–321
1952	90		5	85+	SAMP (Year 1952)	26	196–203
1953	90		5	90+	SAMP (Years 1953–1955)		291–294
1954	90		5	90+	SAMP (Years 1953–1955)		360–363
1955	90		5	90+	SAMP (Years 1953–1955)		434–437
1956	90		5	90+	SAMP (Years 1956–1959)	II	104–115
1957	90		5	90+	SAMP (Years 1956–1959)	II	272–283
1958	90		5	90+	SAMP (Years 1956–1959)	II	438–449
1959	90		5	90+	SAMP (Years 1956–1959)	II	608–619
1960	90		5	90+	SAMP (Years 1960–1962)	II	134–145
1961	90		5	90+	SAMP (Years 1960–1962)	II	364–375
1962	90		5	90+	SAMP (Years 1960–1962)	II	594–605
1963	90		5	90+	SAMP (Years 1963–1964)	II	140–145
1964	90		5	90+	SAMP (Years 1963–1964)	II	312–317
1965	90		5	90+	SAMP (Years 1965–1966)	II	156–165
1966	90		5	90+	SAMP (Years 1965–1966)	II	360–369
1967	90		10	75+	SCD (Years 1966–1967)		210–211
1968–1997	95		1	125+	Detailled Files INSEE (*)		
1998–2014	95		1	125+	www.insee.fr (Detailled Files)		

Notes: SAMP : *Statistique Annuelle du Mouvement de la Population* ; SCD : *Statistique des Causes de Décès*
(*) Detailled Files obtained with ADISP
(2) *Moselle - Bas Rhin - Haut Rhin*
INSEE.fr: `https://www.insee.fr/fr/statistiques/2408054?sommaire=2117120`

Table 5.9: Sources for births, 1901–1935

Year	Départements		Publication	Book	Page
	Total	Missing			
1901	87	(2)	SAMP (Year 1901)	31	32
1902	87	(2)	SAMP (Year 1902)	32	31
1903	87	(2)	SAMP (Year 1903)	33	32
1904	87	(2)	SAMP (Year 1904)	34	32
1905	87	(2)	SAMP (Years 1905 et 1906)	35-36	32
1906	87	(2)	SAMP (Years 1905 et 1906)	35-36	113
1907	87	(2)	SAMP (Years 1907-1908-1909-1910)	1	128–131
1908	87	(2)	SAMP (Years 1907-1908-1909-1910)	1	132–135
1909	87	(2)	SAMP (Years 1907-1908-1909-1910)	1	136–139
1910	87	(2)	SAMP (Years 1907-1908-1909-1910)	1	140–143
1911	87	(2)	SAMP (Years 1911-1912-1913)	2	104–107
1912	87	(2)	SAMP (Years 1911-1912-1913)	2	108–111
1913	87	(2)	SAMP (Years 1911-1912-1913)	2	112–115
1914	77	(1)	SAMP (Years 1914–1919)	3	44–47
1915	77	(1)	SAMP (Years 1914–1919)	3	48–51
1916	77	(1)	SAMP (Years 1914–1919)	3	52–55
1917	77	(1)	SAMP (Years 1914–1919)	3	56–59
1918	77	(1)	SAMP (Years 1914–1919)	3	60–63
1919	90		SAMP (Years 1914–1919)	3	64–67
1920	90		SAMP (Years 1920–1924)	4	34–37
1921	90		SAMP (Years 1920–1924)	4	38–41
1922	90		SAMP (Years 1920–1924)	4	42–45
1923	90		SAMP (Years 1920–1924)	4	46–49
1924	90		SAMP (Years 1920–1924)	4	50–53
1925	90		SAMP (Year 1925) - CD	5	12–15
1926	90		SAMP (Year 1926) - CD	6	12–15
1927	90		SAMP (Year 1927) - CD	7	14–17
1928	90		SAMP (Year 1928) - CD	8	14–17
1929	90		SAMP (Year 1929) - CD	9	16–19
1930	90		SAMP (Year 1930) - CD	10	16–19
1931	90		SAMP (Year 1931) - CD	11	16–19
1932	90		SAMP (Year 1932) - CD	12	14–17
1933	90		SAMP (Year 1933) - CD	13	14–17
1934	90		SAMP (Year 1934) - CD	14	14–17
1935	90		SAMP (Year 1935) - CD	15	14–17

Notes: SAMP: *Statistique Annuelle du Mouvement de la Population*; CD: *Causes de décès.*

(1) *Aisne - Ardennes - Marne - Meurthe et Moselle - Meuse - Moselle - Nord-Oise - Pas de Calais - Bas Rhin - Haut Rhin - Somme - Vosges*
(2) *Moselle - Bas Rhin - Haut Rhin*

Table 5.10: Sources for births, 1936–1971

Year	Départements		Publication	Book	Page
	Total	Missing			
1936	90		SAMP (Year 1936) - CD	16	14–17
1937	90		SAMP (Year 1937)	17	14–17
1938	90		SAMP (Year 1938)	18	114–117
1939	87	(2)	SAMP (Years 1939–1942)	19	78–81
1940	87	(2)	SAMP (Years 1939–1942)	19	138–141
1941	87	(2)	SAMP (Years 1939–1942)	19	200–203
1942	87	(2)	SAMP (Years 1939–1942)	19	260–263
1943	86	Corse + (2)	SAMP (Year 1943)	20	18–21
1944	86	Corse + (2)	SAMP (Year 1944)	21	18–21
1945	87	(2)	SAMP (Year 1945)	22	20–23
1946	90		SAMP (Years 1946–1947)	23	74–77
1947	90		SAMP (Years 1946–1947)	23	132–135
1948	90		SAMP (Years 1948–1949)	24	198–201
1949	90		SAMP (Years 1948–1949)	24	266–269
1950	90		SAMP (Years 1950–1951)	25	196–199
1951	90		SAMP (Years 1950–1951)	25	268–271
1952	90		SAMP (Year 1952)	26	152–155
1953	90		SAMP (Years 1953–1955)		274
1954	90		SAMP (Years 1953–1955)		334
1955	90		SAMP (Years 1953–1955)		408
1956	90		SAMP (Years 1956–1959)	II	53–54
1957	90		SAMP (Years 1956–1959)	II	203–204
1958	90		SAMP (Years 1956–1959)	II	371–372
1959	90		SAMP (Years 1956–1959)	II	541–542
1960	90		SAMP (Years 1960–1962)	II	56–57
1961	90		SAMP (Years 1960–1962)	II	252–253
1962	90		SAMP (Years 1960–1962)	II	494–495
1963	90		SAMP (Years 1963–1964)	II	70–72
1964	90		SAMP (Years 1963–1964)	II	240–243
1965	90		SAMP (Years 1965–1966)	II	69–71
1966	90		SAMP (Years 1965–1966)	II	267–269
1967	90		AS 1968 Tableau XVIII (*)		50
1968	95		SAMP (Year 1968)		136–137; 144–145
1969	95		SAMP (Year 1969)		136–137; 144–145
1970	95		SAMP (Year 1970)		138–139; 146–147
1971	95		SAMP (Year 1971)		140–141; 146–147

Notes: SAMP: *Statistique Annuelle du Mouvement de la Population*; CD: *Causes de Décès*; AS: *Annuaire Statistique*

(2) *Moselle - Bas Rhin - Haut Rhin*

(*) Sinces SAMP in 1967 does not exist, I collect the births for the two sexes and distribute them between boys and girls pro rata births in 1966.

Table 5.11: SOURCES FOR BIRTHS, 1972–2014

Year	Départements		Missing	Book	Page
	Total	Missing			
1972	95		SAMP (Year 1972)		138–139; 148–149
1973	95		SAMP (Year 1973)		138–139; 144–145
1974	95		SAMP (Year 1974)		136–137; 144–145
1975	95		SAMP (Year 1975)		148–151
1976	95		SAMP (Year 1976)		148–151
1977	95		SAMP(Year 1977)		148–151
1978	95		SCD (1978)	II	29–32
1979	95		SAMP (Year 1979)		146–149
1980	95		SAMP (Year 1980)		146–149
1981	95		SAMP (Year 1981)		146–149
1982	95		Collec. de l'INSEE Série D - La Sit. Dem. (Year 1982)		171–174
1983	95		Collec. de l'INSEE Série D - La Sit. Dem. (Year 1983)		171–174
1984	95		Collec. de l'INSEE Série D - La Sit. Dem. (Year 1984)		181–184
1985	95		Collec. de l'INSEE Série D - La Sit. Dem. (Year 1985)		172–175
1986	95		Collec. de l'INSEE Série D - La Sit. Dem. (Year 1986)		172–175
1987	95		Collec. de l'INSEE Série D - La Sit. Dem. (Year 1987)		150–153
1988	95		IR-DS n° 3–4		176–179
1989	95		IR-DS n° 10		174–177
1990	95		IR-DS n° 16–17		212–215
1991	95		IR-DS n° 26–27		186–189
1992	95		IR-DS n° 42–43		188–191
1993	95		IR-DS n° 49–50		188–191
1994	95		IR-DS n° 51–52		188–191
1995	95		IR-DS n° 65–66		188–191
1996	95		IR-DS n° 70–71		217–220
1997	95		IR-DS n° 75–76		194–197
1998–2014	95		www.insee.fr		

Notes: SAMP: *Statistique Annuelle du Mouvement de la Population*; AS: *Annuaire Statistique*; SCD: *Statistique des Causes de Décès*; IR-DS: *Insee Résultats-Demographie et Société*

www.insee.fr: https://www.insee.fr/fr/statistiques/2408051?sommaire=2117120

Table 5.12: Sources for censuses, 1901–2014

Date	Departements		Publication	Book	Table	Variable	Ceiling age
	Total	Missing					
March 4th, 1901	87	(1)	Stat. du RGP Stat.	I to III	I et III	Y. of birth	95
March 6th, 1906	87	(1)	Stat. du RGP Stat.	II and III	II	Y. of birth	80
March 5th, 1911	87	(1)	Stat. du RGP Stat.	II	III	Y. of birth	105
March 6th, 1921	90		Stat. du RGP Stat.	II and III	I	Y. of birth	80
March 7th, 1926	90		Stat. du RGP Stat.	II and III	I	Y. of birth	80
March 8th, 1931	90		Stat. du RGP Stat.	II and III	I	Y. of birth	80
March 8th, 1936	90		Stat. du RGP Stat.	II and III	I	Y. of birth	80
March 10th, 1946	90		RGP - Results by dept	I to VI	I	Y. of birth	80
May 8th, 1954	90		RGP - Results by dept	I to VI	D1	Y. of birth	89
March 7th, 1962	94	(2)	DE - Results by dept	I to VI	D1	Y. of birth	84
March 1st, 1968	95		www.insee.fr			Age	120
Feb. 20th, 1975	95		www.insee.fr			Age	120
March 4th, 1982	95		www.insee.fr			Age	120
March 5th, 1990	95		www.insee.fr			Age	120
March 8th, 1999	95		www.insee.fr			Age	120
January 1st, 2008	95		www.insee.fr			Age	120
January 1st, 2013	95		www.insee.fr			Age	120
January 1st, 2014	95		www.insee.fr			Âge	120

Notes: RGP: *Recensement Général de la Population*; DE: *Dépoullement Exhaustif*

(1) *Moselle - Bas Rhin - Haut Rhin*

(2) In 1962, the census made in *Corse* was irrelevant (cf p. 5 of the book). Only the 1/20th exploitation available in the regional *Provence-Alpes-Côte d'Azur* book was used. This one provided population by quinquennial group of birth years while the last class provided the 74 year-old and over, not the 84 year-old and over. To compute these age classes and get an homogeneous census, I use the distribution of the other *départements*. As an exemple, for ladies born between 1958 and 1962, 23.95% were born in 1961 elsewhere. So I apply this percentage on the sum of ladies born between 1958 and 1962 in *Corse* (4,860) and I deduct that 1,164 were born in 1961.

www.insee.fr : `https://www.insee.fr/fr/statistiques/2414232`

Chapitre 6

Computations of French Income Distributions by *Département*, 1960–2014

Abstract

Creating a unique database on spatial inequaliy in the long run, we shed new light on the recent debates on French territorial divide. This paper presents the methods we used to compute income distributions of each French metropolitan département during the period 1960–2014. We first present the raw materials that we collected in the archives of the Finance Ministry and which cover the periods 1960–1969, 1986–1998 and 2001–2014. Then we present our method to compute income distributions from this raw material. We built on the methodology developped by Piketty (2001) for France and extended by Garbinti et al. (2016) and Blanchet et al. (2017). It relies on the Pareto law and its extrapolation. Using this new and unique dataset, we explore the evolution of the departmental income distributions for the last 50 years. In particular, we show that the average income per département has converged whereas top 1% tax units are more and more unevenly located.[1]

[1]This paper is co-written with Aurélie Sotura.

6.1 Introduction

France, unlike many of its European neighbors, is a unified country whose metropolitan borders haven't changed much for the last 150 years.[2] This old administrative unification does not imply economic unification : France hides strong territorial disparities, and debates on the issue of the territorial divide are numerous. As such, Labrador (2013) estimates that the median of monthly fiscal income was €1,257 per Consumption Unit in *Seine-Saint-Denis* in 2010, compared to €2,087 in Paris. These differences observed in the Paris metropolis are also strong between Paris and the countryside.

Although these spatial inequalities are relatively well known over the recent period, little is known of their long-term dynamics. This paper looks at spatial inequalities of income in a historical perspective. As such, we have estimated the income distributions of all *départements* in metropolitan France since 1960 using archival documents that had never been exploited, and Blanchet et al. (2017)'s methodological protocol that makes reference. As a result, we go further than the analyzes of average income per geographic unit. Even if this kind of analysis gives an idea of the spatial inequalities, they hide a big part of the territorial differences by considering that intra-departmental inequalities are non-existent. With this new database, we know where the poorest or the richest households are located on the national territory, for more than 50 years. This is a major breakthrough in the knowledge of French spatial development since some of the results we observe today are explained by path dependence : the decisions made yesterday have long-lasting effects.

This work has never been done for several reasons. The first is that the amount of raw material from the archives is considerable. We have digitized more than 4,500 fiscal tabulations for this paper. Archives used are administrative documents never officially released. This partly explains why most of the studies on the issue of income inequality have focused on income distributions at the national level. According to France, Piketty (2001) was the first to study in detail income inequalities during the 20th century. For this purpose, he used fiscal tabulations collected by the administration and available in official publications as his main statistical source. Among the numerous results of his work, he showed that the top percentile share of income fell sharply between 1900 and 1946, from 20% to 6%. This share remained stable until 1998, when he estimated it at around 8%. These pioneering works were later extended by Garbinti et al. (2016). This paper consists mainly of extending data from 1998 to 2014, getting estimation per individual and not only per tax unit, and getting series after and before redistribution. Moreover, Piketty (2001) only estimated the income shares of the first decile while Garbinti et al. (2016) used a new method in order to compute the entire distribution. They relied on Blanchet et al. (2017). The work done for France gave birth to a vast literature in this field : methods were replicated in countries where the statistical sources allowed it.[3] Today, an

[2]The last major change dates back to the incorporation of *Savoie* and Nice's *comté* in 1860. Subsequently, only *Alsace-Moselle départements* have escaped temporarily from the French administration between 1870 and 1918 and during the Second World War.

[3]For example, the works of Atkinson (2005) for the United Kingdom, Roine and Waldenström (2008) for Sweden, Atkinson and Salverda (2005) for the Netherlands, Alvaredo and Saez (2009) for Spain or Alvarado (2009) for Portugal can be cited.

online database[4] provides these inequality indices as well as estimation methods and software codes.

Second, up to now, the literature on economic history has focused on value added rather than income. A vast literature emerged following Geary and Stark (2002), who estimated Ireland and United Kingdom's value added at the subnational level over a very long period. Their method is based on the joint knowledge of values added by sector at national level and of employment and wages for each region. Assuming that wages are good approximations for local productivities, the authors allocate national value added between regions. Due to the parsimony of data needed to make these estimates, this method has been used to estimate regional value added of a large number of European countries over a long period.[5] Each of these works shed special light on inter-regional inequalities. For example, Buyst (2010) showed how the relative regional positions were reversed during the 20th century in Belgium : Flanders became the richest region while Wallonia, once rich in natural resources, became the poorest one. Felice and Vecchi (2015) showed how Italy's north and center gradually converged, leaving the *Mezzogiorno* more and more isolated.

For France, the lack of reliable data on local wages has prevented an application of this method. In fact, regional value added estimates are more fragmented. However, three works can be cited in this field. The most important is the one of Combes et al. (2011). This paper built departmental value added for three dates (1860, 1930, 2000) and studied the main trends in the evolution of spatial inequalities, both in terms of population and sectoral value added. The second is the one of Bazot (2014), which built departmental value added between 1840 and 1911, for every ten years. It uses tax statistics, and in particular those relating to the *patente*, an old tax based on the non-agricultural production. Lastly, one can mention Caruana-Galizia (2013) who develops an econometric model based on sectoral employment in order to estimate departmental value added between 1871 and 1911.

Finally, our study is at the frontier of these two fields, namely income distributions at the national level, and spatial inequalities of value added within countries. We are in line with Sommeiller and Prize (2014) and Sommeiller et al. (2016), who were the first to apply the methods used by Piketty and his colleagues to estimate subnational income distributions (namely in the states of USA since the early 20th century). We contribute to this emerging field by producing these data within the framework of the French metropolitan *départements*.[6] Although this paper does not aim to analyze the process of convergence within French *départements*, the first results show that spatial inequalities have decreased since 1960 according to income per adult. At the same time, spatial inequalities have increased in specific regions : this is particularly true in *Ile-de-France*, where the homogeneous situation of the mid-1960s contrasts with the strong spatial segregation visible nowadays. Finally, we show a gentrification process in some *départements* : the poorest tax units are gradually disappearing from the territories close to the Swiss border (namely *Haute-Savoie, Ain* and *Bas-Rhin*).

[4] `http://wid.world/fr/accueil/`

[5] We can cite the works of Felice and Vecchi (2015) for Italy between 1860 and 2010, Badia-Miro et al. (2012) for Portugal between 1890 and 1980, Buyst (2010) for Belgium between 1896 and 2000, Enflo and Rosés (2015) for Sweden throughout the twentieth century or Schulze (2007) for the former Habsburg empire between 1870 and 1910.

[6] The French overseas *départements* are treated in another study. The main results in the case of *La Réunion* are available in Govind (2017).

The rest of this paper is as follows. In Section 2 we present data used to compute income distributions. In Section 3 we present the methods used to make these estimates, which differ according to the two sub-periods 1960–1969 and 1986–2014. Section 4 present some of the results available. The fifth section concludes.

6.2 Data and Spatial Framework

6.2.1 Fiscal Data at the Departmental Level

Fiscal data come from a single administrative document, namely *"Etats 1921"*, established by the fiscal administration. These documents, later called "fiscal tabulations", document the number of fiscal units and the total of fiscal income for each income brackets previously defined by administration. These income brackets vary in time, as the income tax schedule changes. Piketty (2001) was the first to make a systematic use of these documents at the national level over the period 1915–1998, in order to estimate inequalities between French tax units.

It turns out that these documents were also produced for each metropolitan *département*, even if they were not published in official publications. Nevertheless, they are freely available in the Finance Ministry's archives for the two periods 1960–1969 and 1986–1998. For the period 2001–2014, these documents are freely available online. These statistical documents have two main advantages for our study. The first is that they are available in the Paris region for all metropolitan *départements*, and full of details.[7] The second is that they concern income tax, which applies uniformly throughout the territory. Consequently, the methodology used at national level by Piketty (2001) and then by Garbinti et al. (2016) can be globally replicated at the departmental level.

Data are not of the same quality depending on whether one is interested in the period 1960–1969 or in the period 1986–2014. For the former one, fiscal tabulations of taxable and non-taxable taxpayers are available. We have recovered both of them. No fiscal unit is supposed to be missing in these data. This is not the case between 1960 and 1969 : only tabulations for taxable fiscal units are available. We will present in detail assumptions made to estimate the non-taxable fiscal units for this period.

6.2.2 Departmental Demographic Data

Departmental demographic data are of crucial importance in this study. Indeed, we have to estimate the number of tax units for each *département* between 1960 and 1969. As such, we use the departmental demographic database built by Bonnet (2018), which covers the entire 20th century. This paper used both raw population-by-age data for each census and yearly age-specific deaths. With these data he estimated populations by age on each January 1^{st}. We have extracted populations by age and *département* between 1960 and 2014.

In addition, we have collected census data on the number of married couples in each *département*. These statistics are available in 1962 census books, then on the INSEE website.[8]

[7] We only use part of these documents for this paper. They contain tabulations by category of income and by category of household (demographic structure).

[8] These documents have been downloaded on April 30th, 2017 on : `https://www.insee.fr/fr/statistics/`

6.2.3 National Data

We have collected fiscal tabulations established at the national level. However, they do not concern the same spatial frame as they include the Overseas *départements* (i.e. *Guadeloupe*, *Martinique*, *Guyan*e, *Réunion*, as well as *Mayotte* since 2015). Given population weight of these territories (around 3% of the national population in 2014), Garbinti et al. (2016)'s indicators calculated from these tabulations allow us to check the consistency of our estimates. We also retrieved national accounts data, including estimates of taxable income, fiscal income and gross primary income, available in Garbinti et al. (2016)'s appendices. Finally, we retrieved the distribution of taxable income thresholds and enhancement rates necessary to transform the distribution of taxable income into a distribution of fiscal income. These enhancement rates are available in Garbinti et al. (2016)'s appendices too.

6.2.4 Spatial Framework

The administrative unit used for this study is the *département*. *Départements* have an advantage compared to other geographical divisions of the French metropolitan territory : their administrative boundaries have remained largely stable since their creation in 1789. Nevertheless, we have to deal with one change in administrative boundaries : the law promulgated on July 10, 1964 reorganized *Ile-de-France*. Until that year, this region was composed of three *département*s : *Seine, Seine-et-Oise* and *Seine-et-Marne*. *Seine-et-Marne* remained the same after this date, but it was not the case of *Seine* and *Seine-et-Oise*. Seven new *départements* have come out : *Essonne, Hauts-de-Seine, Seine-Saint-Denis, Val-de-Marne, Val d'Oise, Paris* and *Yvelines*. Administratively, this transition was fixed at January 1, 1968. However, in our statistics, classification changes in 1966. Since the old *départements* are not a linear combination of the new ones, we cannot estimate directly departmental distributions according to the old classification. As such, our estimates relate to the old classification up to 1965, and to the new one from 1966 onwards.

6.3 Methods

6.3.1 Period 1986–2014

6.3.1.1 Raw Fiscal Statistics Available

Fiscal tabulations of the period 1986–2014 include all tax units, whether taxable or not.[9] The main problem is the temporal consistency of the taxable income brackets used by the tax administration. These brackets are important for our computations. As Blanchet et al. (2017) explain, the lower threshold of the highest bracket should not be too low to accurately estimate high incomes, while the upper threshold of the lowest bracket should not be too high to estimate in a proper manner the distribution of income for the poorest. In addition, the total number of brackets should not be too

2414232.

[9]In the raw materials we have collected, each *département* has two different tabulations : the first for taxable units only, the second for tax-free units. We have summed these two tabulations.

small to accurately estimate the distribution's core. Table 6.3 in Appendix shows how the thresholds have evolved since 1986. The first observation is a perfect stability of these brackets between 1986 and 1998. There are twelve successive brackets, from FF40,000 (*francs*) to FF500,000 and more taxable income per tax unit. Despite the upward trend of these taxable incomes due to growth and inflation, this stability and the large number of brackets are optimal for our estimates. However, the situation is different since 2001. Tax instability leads to numerous threshold changes. The number of brackets is particularly low in 2001 and 2002.

In order to improve computations of these two specific years, we take advantage of two specific elements. First, the brackets are the same as the ones used between 2003 and 2005, but with less detail. Thus, the €0–9,000 bracket in 2001-2002 is divided into two brackets in 2003 : €0–7,500 and €7,500–9,000. Second, the national fiscal tabulations according to 2001 and 2002 exist for the same decomposition as in 2003. We decide therefore to use these two pieces of information to refine the bracket values in 2001 and 2002. Let call p_{ij} the share, for the department i, of the sub-bracket j in the global bracket. We assume :

$$\left(\frac{p_{ij}}{p_{natj}}\right)_t = \left(\frac{p_{ij}}{p_{natj}}\right)_{2003} \text{ for } t = \{2001,\, 2002\}\,. \tag{6.3.1}$$

In other words, we consider that the weights of each sub-bracket at the departmental level vary similarly from what is observed at the national level. Since national values are known for each year and for 2003 at the departmental level, we deduce the sub-bracket size for 2001 and 2002 years.

At the end, management of raw fiscal tabulations causes some issues because of the small numbers in the rural *départements*. Thus, when intermediate brackets do not contain tax units, we sum the tax units and the taxable income of all higher brackets. However, this operation was done only for very isolated cases.

6.3.1.2 Taxable Income and Fiscal Income Distributions

Income as defined in fiscal tabulations, so-called "taxable income", is not the economic aggregate on which we have based our estimates. In fact, taxable income is defined as the reported income from which a set of allowances has been deducted. These allowances changed over time as the tax legislation changed too. The "fiscal income" we use is net of deductions from all these allowances. Moreover, one has to use carefully taxable income as soon as it sometimes includes capital gains and sometimes not. Capital gains are the gains realized on stock markets. They are mainly exceptional and are not included in the same way in tax tabulations as are recurrent incomes. Yet, they provide a useful and complementary perspective in understanding inequality. In their study, Garbinti et al. (2016) compute distributions of fiscal income based on two specifications,with or without capital gains. We decide to do similarly.

With the data we get in archives we compute in a first stage a taxable income distribution for all tax units.[10] In a second stage we transform the distribution of taxable income into a distribution of fiscal income. One has to remember that we want to get two kind of fiscal income distributions : one with

[10]For this, we used the *gpinter* package provided by Blanchet et al. (2017), which extrapolates the entire taxable income distribution from fiscal tabulations only.

capital gains, and another that does not take them into account. To do so, we use the enhancement rate profiles calculated by Garbinti et al. (2016) at the national level. These enhancement rates are different according to the fractile, and vary strongly within the first decile of income. In addition, these enhancement rates are different depending on whether the taxable income incorporate capital gains or not. This point is crucial as soon as our raw statistics are not homogeneous. The period 1986–1998 does not include them while they are included between 2001 and 2014. We apply national enhancement rates differently across the two sub-periods to provide consistent estimates over time.

In a third stage we reconcile the sum of the fiscal income established at the departmental level with the fiscal income at the metropolitan level. This is important as we want our estimates perfectly consistent with those of Garbinti et al. (2016). We proceed with a general adjustment so that the two sums correspond.[11] We start from the national fiscal income, very close to the metropolitan fiscal income (difference due to Overseas *départements*), and compute the metropolitan fiscal income (FI^{met}) as follows :

$$FI_t^{met} = TI_t^{nat} \times \frac{FI_t^{nat}}{TI_t^{met}} \quad \text{for } t = \{1986, ..., 1998, 2001, ..., 2014\}, \tag{6.3.2}$$

with TI^{nat} and FI^{nat} the taxable and fiscal incomes at the national level, and TI^{met} the metropolitan fiscal income.

Figure 6.3.1 shows the fiscal income shares held by the top 0.1% tax units for each specification, at the metropolitan level by summing the departmental tabulations, and at the national level. For each of the specifications, the shares of the 0.1% richest fiscal units at metropolitan and national level are very close, regardless of the specification. One important difference remains between 2001 and 2002 : the share deviates 0.2 percentage points from the national level.

6.3.2 Period 1960–1969

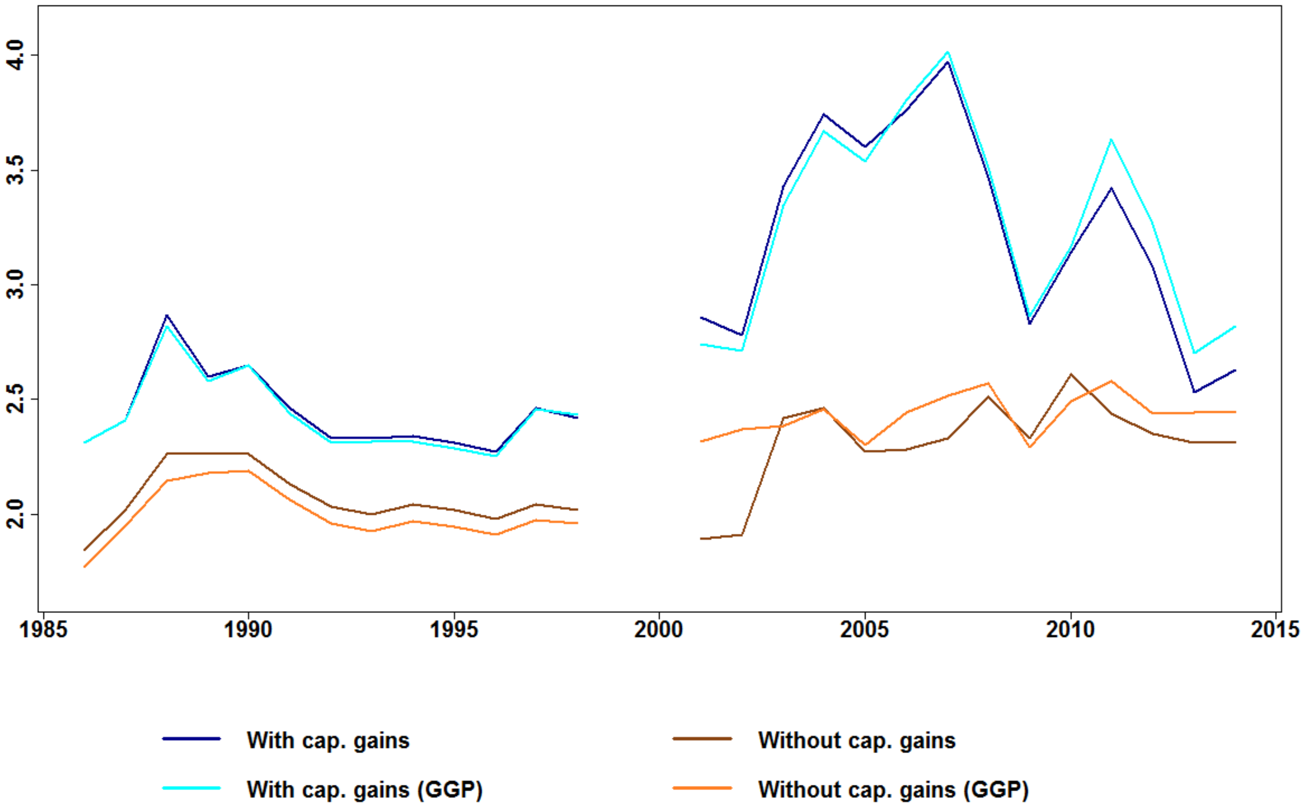
The period 1960–1969 presents two main issues for the estimation of the departmental distributions of income. The first is that fiscal tabulations only concern taxable units. To tackle this problem we estimate the total of tax units in each *département*. The second is linked to the first : since tax tabulations do not allow to compute the total of fiscal income in each *département* as we do for the period between 1986 and 2014, we estimate it with a different method.

6.3.2.1 Computations of Total Tax Units by *Département*

We test two methods to estimate the total number of fiscal units for each *département* and each year for the period 1960–1969. The first comes from Govind (2017), who assumed that the sum of fiscal units is equal to the population of more than 20 years less the married couples (who have to fill out a single tax return). She considered that all individuals over age 20 complete a tax return. However, this age changes over the years, and is not the same for all *départements*. The second method is developed specifically for this study. It is based on the period 1986–2014 for which we get the total

[11] The difference between the metropolitan fiscal income and the sum of the departmental fiscal income is low (globally 1% for each year).

Figure 6.3.1 : Top 0.1% share according to the specification



Notes : GGP means Garbinti et al. (2016)'s work. "With capital gains" means that capital gains are included in fiscal income. Sample includes 95 *départements* in our specification, 99 *départements* in Garbinti et al. (2016).

number of fiscal units for each *département* i and each year t in fiscal tabulations. We estimate these values as follows :

$$FH_{it} = \alpha_0 + \sum_{a=1}^{6} \alpha_a P_{it}^a + \beta M_{it} + \gamma FH_{natt} + \delta_i + \varepsilon_{it} \qquad (6.3.3)$$

FH_{nat} is the total of tax units at the national level coming from Garbinti et al. (2016), M is married couples and P^a is the population of age a. In addition, we have summed the population by age into six major age groups and we have decomposed the error into an error specific to each *département* δ_i and an econometric error ε_{it}.

Table 6.1 presents the distribution of the differences (in %) between departmental fiscal units estimated and values observed for the years 1986 to 2010.[12] On average, the first method estimates tax units properly between 1995 and 2010. However, one can view some errors of more than 5% (Columns 6 and 10). For the previous period, values are overestimated : Table 6.1 reveals that 50% of the *départements* have an error greater than 2% in 1990. For 25% ot them, this error is greater than 4.5% (Column 7).

[12] We do not use the years 2011 to 2014 as it presents a structural change with the previous period. This is due to a change in tax legislation concerning the definition of the tax unit : until that date, married and divorced couples completed three returns in the year of their union or divorce : one for each spouse, and one for the couple, which is no longer true.

The method we develop seems more efficient in estimating the total of tax units in the metropolitan *départements*. Overall, the R^2 of the model is equal to 0.9996. Systematic errors do not exist : the annual median of differences is always between -0.7% and 0.7% (Column 3). However, some important errors are visible punctually, especially in 1986. In detail, *Corse* seems to be problematic : the model greatly overestimates the number of tax households (about 20% in 1986). This error may be due to the specificity of this *département* at the time. However, it can came from the difficulties of the fiscal administration in this *département* too. Beyond this specific example, 1986 is the first year during which tax units are enforced to declare their income, regardless of whether they are taxable or not. Some of them may have not done so since they were not used to these formalities.

Table 6.1 : Differences between tax units recorded and estimated

Method	Bonnet-Sotura					Govind				
	Min.	Quart. 1	Méd.	Quart. 3	Max.	Min.	Quart. 1	Méd.	Quart. 3	Max.
1986	-3.6	-0.8	0.5	3.1	20.1	-0.2	2.65	4.7	9.95	46
1990	-2.9	-1.35	-0.7	0.3	6.1	-1.5	0.75	2	4.5	24.2
1994	-2.4	-0.3	0.1	0.6	2.9	-1.3	1.1	1.8	3.4	17.2
1998	-3.8	-0.8	-0.2	0.25	1.7	-4.4	-1.1	-0.3	0.9	6.3
2002	-4	-0.55	0.3	0.65	1.7	-7.3	-1.7	-0.6	0.25	4.2
2006	-3.8	-0.4	0	0.65	3.1	-7.4	-2.35	-1.4	-0.7	3.2
2010	-3	-0.2	0.7	1.45	5	-7	-2.45	-1.2	-0.2	4.1

Notes : Differences in %. Bonnet-Sotura estimate tax units by an econometric method based on demographic data. Govind estimate tax units by substracting married couples from adult population. "Quart. 1" means that 25% of observations have a difference lower than the threshold. "Med." is the usual median. Sample includes 95 *départements*. Values in % are ratio between tax units estimated and recorded.

To verify whether these estimates are consistent or not, we sum up the estimated departmental tax units and compared them to the national tax units compiled by Garbinti et al. (2016) between 1960 and 2014. Firstly, national tabulations reveal a total of tax units 2 to 3% higher than metropolitan tax units between 1986 and 2014, with an increasing difference over time. This is the weight of overseas *départements*. Second, the model faithfully follows this gap, except between 2011 and 2014 and in 1986. Years 2011 to 2014 can be explained by the fall of the total number of tax units in the tabulations, due to the change in law. Third, the sum of departmental tax units is overestimated between 1960 and 1969. If we consider that the weight of overseas *départements* is 1% over the period, this overestimation is around 5 points in 1960, 2 points in 1965 and 1.25 point in 1969. We take this into account by applying a uniform adjustment coefficient for the period 1960–1969. With this adjustment, national tax units are equal to 101% of metropolitan tax units.[13]

[13] An other specification using year-fixed effects instead of department-fixed effects has been tested. This specification shows lower R^2, but gives roughly the same results after the uniform adjustment. Since we are not interested in the values of the coefficients of the variables but in the values of fiscal households estimated, we have decided to present only this specification.

6.3.2.2 Computations of Fiscal Income by *Département*

Based on the fiscal tabulations for taxable tax units, the distributions of fiscal income can be established similarly as we do between 1986 and 2014 (i.e. by applying the enhancement rates to the taxable income distribution). To calculate the entire distribution of fiscal income, and not the one for only taxable units, we must estimate the fiscal income of each *département*. To do so, several alternatives have been tested. The one that seemed the most relevant relies on the exhaustive knowledge of the departmental distributions for the period 1986–2014.

Using fiscal tabulations – which give us the number of taxable tax units – and the estimation of the departmental tax units as presented in Section 6.3.2.1, we get the share of tax-free units in the total, by *département* and year of the period 1960–1969. We must estimate the share of these tax-free units in the total of fiscal income for each *département*. We do so by using the departmental Lorenz curves between 1986 and 2014. These curves are modeled as follows :

$$L_{xt} = \alpha_0 + \alpha_1 Sh_x + \alpha_2 (Sh_x)^2 + \alpha_3 L_{xt}^{nat} + \delta_x + \varepsilon_{xt}. \tag{6.3.4}$$

For each *département*, L_x is the share of income held by the x% of the richest tax units in year t and Sh_x is the share of the x% richest tax units. L_{xt}^{nat} is the income share held by the x% richest tax units in year t at the national level, which comes from Garbinti et al. (2016). We include a fixed effect specific to quantile x, δ_x. ε_{xt} is an econometric error term. In other words, we use the information coming from 1986–2014 Lorenz curves at the departmental level and the information coming from 1960–2014 Lorenz curves at the national level in order to estimate the 1960–1969 Lorenz curves at the departmental level. The R^2 of Equation (6.3.4) is equal to 0.995, quite similar according to the specification used (i.e. with or without capital gains).

Finally, we get values of $\hat{L_{xt}}$ and thereby the share of tax-free units in fiscal income. We assume that tax-free units are tax units with the lowest incomes.[14] Finally, we call FI^{tax} the fiscal income of taxable units and $\hat{L\bar{tax}}$ the share of tax-free units in the fiscal income. We obtain the total fiscal income as follows :

$$FI = \frac{FI^{tax}}{1 - \hat{L\bar{tax}}}.$$

We also tried another method based on regional accounting. Nevertheless, the results were not relevant since fiscal income for tax-free units in *Ile-de-France* was negative. Although we do not use these results, we present this methodology in Appendix 6.6.1.

6.3.2.3 Computations of Fiscal Income Distributions for all Tax Units

With fiscal income distributions of taxable units and fiscal income of tax-free units for each *département*, we finally compute fiscal income distributions for all tax units. We follow the guidelines of Garbinti et al. (2016). We do not use the entire fiscal income distribution of taxable units, but only

[14]This is an approximation since some tax units with high incomes could be tax-free units since they have a high *quotient familial* (mainly numerous children). These situations are however not very widespread.

the distribution above a fiscal income threshold. These annuals thresholds are defined in such a way that all high incomes are included in the distribution, regardless of their *"quotient familial"* (number of adults and children in the fiscal unit).[15]

6.3.3 Template of Fiscal Income Distributions by *Département*

Departmental distributions of income available follow Alvaredo et al. (2016)'s guidelines. The distributions extracted contain, for each fractile, different informations. Firstly the income threshold that must be exceeded to be part of the fractile, then the average income of the fractile, the share of total income owned by the fractile, the total income share held by all tax units above the threshold, the number of tax units in each fractile, as well as the income threshold that must be exceeded to enter the fractile, in % of the average departmental income.

6.4 Results

These departmental distributions of income over a 55-year-period opens the way to a number of specific studies. They should help to better understand how the households are spatially distributed and the reasons why. We have chosen to present in this section some of the results available through our study. They include average fiscal income per adult, intra-departmental inequalities, as well as the spatial distribution of tax units belonging to each fractile. All the results are calculated using the specification without capital gains.[16]

6.4.1 Fiscal Income per Adult

The first contribution of this study lies in the knowledge of the fiscal income of each *département*. As a result, one can deduce the average fiscal income per adult[17], as well as the average fiscal income per adult relative to the metropolitan average. Figure 6.4.1 maps these relative fiscal incomes in 1966 and 2014.[18] The values for the top and last 20 *départements* ranked by average income per adult in 2014 are available in Section 6.6.2 of the Appendix (Table 6.4).

The 1966 map shows that high incomes were concentrated in the Paris region : the fiscal income per adult were equal to 168% of the metropolitan average for Paris and 161% for *Hauts-de-Seine*. The situation was relatively homogeneous for the whole region (excluding *Seine-et-Marne*). The lowest value was in *Seine-Saint-Denis* : nevertheless, the average income per adult was 24% higher than the metropolitan average. There was also a North-East/South-West gradient : the South-West rural *départements* such as *Dordogne*, *Aveyron* or *Ariège* had the lowest values of the panel, around 60-

[15]Due to multiple children, some high-income tax units could not pay income tax and thus be out of the fiscal tabulations. In order not to underestimate inequalities, one should estimate how these households are important and reintegrate them in the distribution. These computations are almost impossible at the departemental level.

[16]Maps of this section have been realized with the software Philcarto, `http://philcarto.free.fr/`.

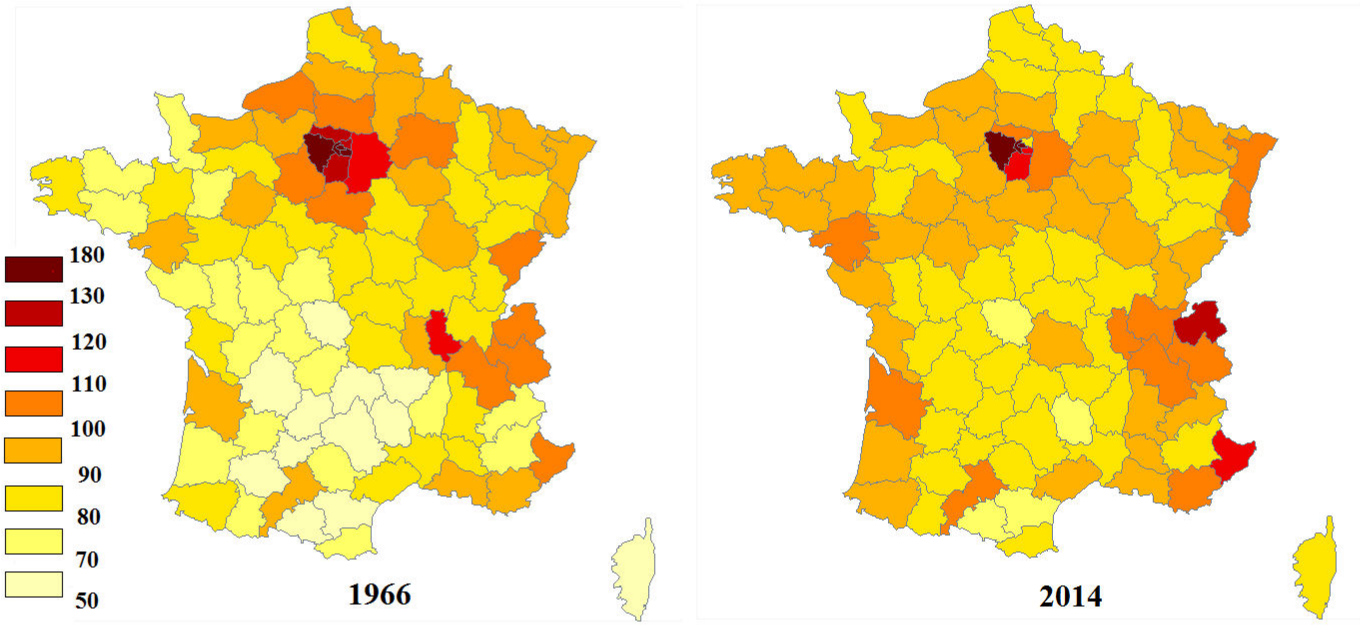
[17]Adults are defined as the population aged 20 and over.

[18]1966 is the first year we get departmental values in the current classification while 2014 is the last year we currently compute

70% of the metropolitan average. The urban *départements* of this region were doing better, namely *Haute-Garonne* (which hosts *Toulouse*) and *Gironde* (which hosts *Bordeaux*).

In 2014, the situation is noticeably different. From a global point of view, the geography is much more homogenized. The lowest values are still in the South-West, but they have increased : the average income per adult is 74% of the metropolitan average for *Creuse* (58% in 1966), and 78% for *Lozère* (59% in 1966). There are three main geographical areas that have changed according to average income per adult. The first is the Paris region, whose homogeneity no longer exists. Paris, *Hauts-de-Seine* and *Yvelines* still have values 40 to 60% above average, but all other *départements* drop. This is the case in *Val-de-Marne*, which is 25 points down (139% in 1966 against 114% in 2014). The situation is also dramatically different in *Seine-Saint-Denis*, whose values are respectively 124% and 84%, a drop of 40 points. It is the only *département* in the *Ile-de-France* region whose average income per adult is lower than the metropolitan average in 2014. The second geographical area is the Atlantic-Brittany one, whose average income has risen sharply for all these *départements*. Their situation is very homogeneous, around 90-100%. Within this area, *Gironde* and *Loire-Atlantique* (formerly *Loire-Inferieure*), stand out with values of 100%. The third geographical area is the Swiss border. *Haut-Rhin, Jura, Ain* and *Haute-Savoie* emerge particularly in 2014. In this respect, *Haute-Savoie* is the fourth highest income-per-adult *département*, with an average income 29% higher than the metropolitan one.

Figure 6.4.1 : Relative fiscal income per adult, 1966 and 2014



Notes : Values are percentage of the metropolitan average. Computations made with the no-capital gains specification. Sample includes 95 *départements* in 1966 and 2014.

6.4.2 Intra-departemental Inequalities

Another contribution of this study lies in the knowledge of tax income inequalities within each *département*. Piketty (2001) analyzed the shares of income at the national level. He showed that the top percentile share declined throughout the 20th century, from 20% to 6% between 1900 and 1946, before stabilizing at around 8% nowadays. Figure 6.4.2 maps the top 10% share of income for each *département*. Section 6.6.2 of the Appendix (Table 6.5) provides top-to-bottom ratios – the ratio of the top 5% average income to the bottom 95% average income – for the first and the last 20 *départements* according to their values in 2014.

Figure 6.4.2 reveals the top 10% share of income in 1966. If we except *Ile-de-France* and four specific *départements*, one can see an homogeneous landscape. In each *département*, these shares fell between 34 and 38%. This means that the average income of these tax units was 3.4–3.8 times higher than the departmental average income. These results are different from the ones presented in Figure 6.4.1 : intra-departmental inequalities were more homogeneous than inequalities between *départements* from a spatial point of view. The Paris region was an exception. The lowest values of all metropolitan *départements* were in *Val-d'Oise, Seine-Saint-Denis, Val-de-Marne* or *Essonne*, while Paris had the highest inequality levels. *Yvelines* and *Hauts-de-Seine* are in an intermediate situation, close to other metropolitan *départements*.

Figure 6.4.2 reveals these values in 2014 too. Compared to 1966, intra-departmental inequalities have decreased. Once again, apart from the Paris region, the landscape is homogeneous : the share of the top 10% tax units fall between 28% and 32% everywhere : average income of these tax units is 2.8–3.2 times higher than the departmental average income. Some exceptions remain : the Mediterranean coast, but also *départements* such as *Haute-Garonne, Gironde, Haute-Savoie, Marne* or *Nor*d. Within these *départements*, inequalities are more important, with values between 32 and 34%. According to the Paris region, the findings made in 1966 are still valid. Paris and *Hauts-de-Seine* have the highest values : in Paris, the top 10% share of income is 44% (38% in *Hauts-de-Seine*). On the other hand, intra-departmental inequalities are lower in *Seine-Saint-Denis, Seine-et-Marne* and *Val-d'Oise*, with values between 30 and 32%.

6.4.3 Spatial Distribution of Tax Units Belonging to each Fractile

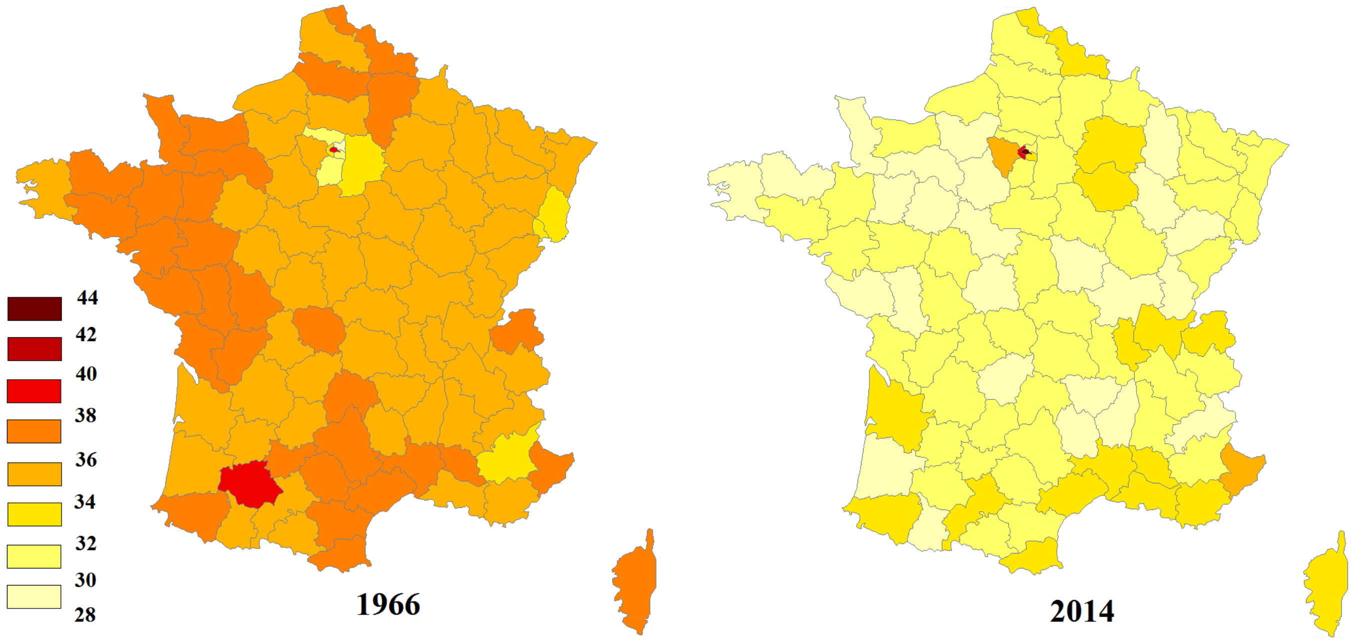
With our new database, we know the fiscal income thresholds one tax unit has to exceed to be part of each fractile defined at the metropolitan level. Then we can study the spatial distribution of the poorest and the richest tax units. However, with absolute values we cannot know whether the poorest or the richest tax units are overrepresented in each *départment*. To do so, we take into account the *département* population share in the metropolitan total. As such, we compute distortion indices defined as follows :

$$ID_i^x = \frac{\frac{P_i^x}{P_i}}{\frac{P_{met}^x}{P_{met}}} = \frac{\theta_i^x}{\theta_{met}^x},$$

where i is the *département,* P^x is the number of tax units belonging to the fractile x and P is the total of tax units. Consequently, θ_i^x is the share of tax units belonging to the decile x in *département* i. Since decile thresholds are set at the metropolitan level, θ_{met}^x values are simple (0.1 for each decile, and 0.01 for each percentile).

Figure 6.6 maps these distorsion indices for the last five deciles, that is the 50% poorest tax units.

FIGURE 6.4.2 : TOP 10% SHARE OF FISCAL INCOME, 1966 AND 2014



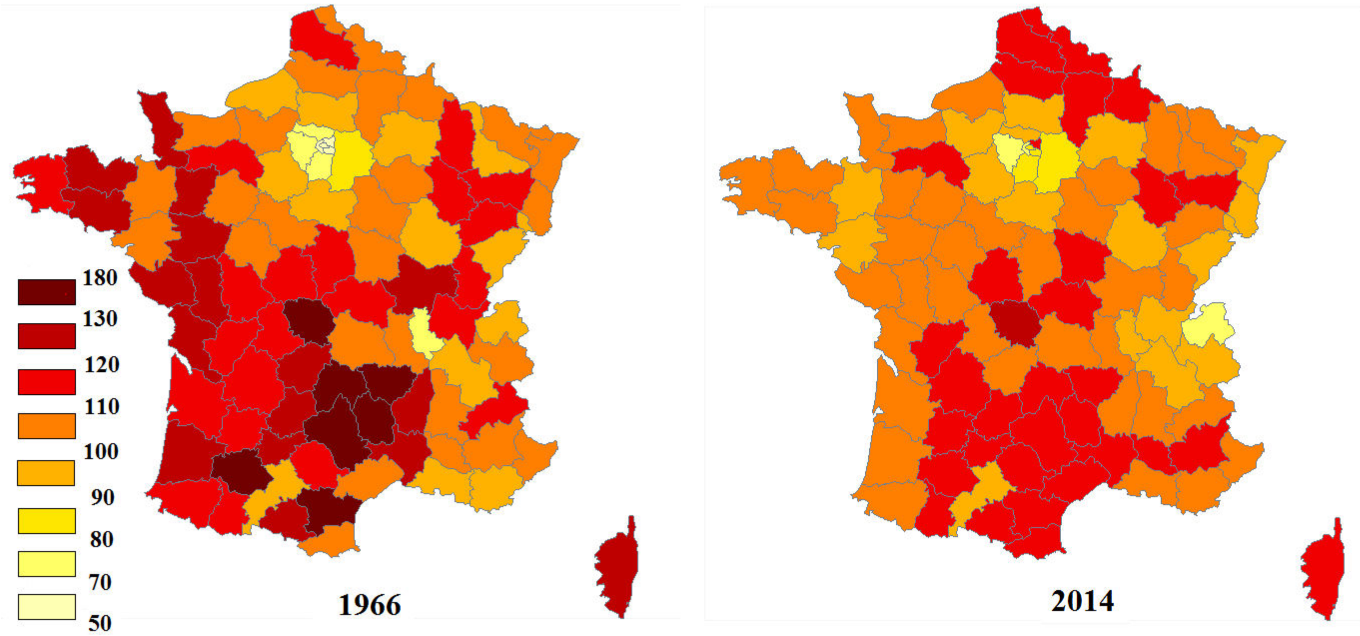
Notes : Values are percentage of the metropolitan average. Computations made with the no-capital gains specification. Sample includes 95 *départements* in 1966 and 2014.

Values for the first and the last 20 *départments* (ranked according to their 2014 values) are available in Section 6.6.2 of the Appendix (Table 6.6). The first map shows the values for the year 1966. One can see that the poorest were overrepresented in the South-West of the country. This is particularly true in the south of *Massif Central*, in *Aveyron*, *Lozère* or *Cantal*. In these *départements*, the share is 30% higher than the metropolitan one. High values were also found in *Bretagne* as well as in *Vosges* and *Meuse*. Conversely, the Paris region and the *Rhône département* had very low values, around 60-70%.

The second map reveals the values in 2014. Overall, the poorest are more evenly distributed in the metropolitan area than in 1966. Only *Creuse* keeps an index higher than 120% (124%). The poorest are still overrepresented in the South-West *départements*, with a homogenous geographical area corresponding to the *Occitanie* region from which *Haute-Garonne* would be withdrawn. In addition, two new geographical areas appear in the geography of poverty. The first is *Seine-Saint-Denis*, whose index has significantly increased in 50 years to 115%. North of France is also an homogeneous geographical area in which the share of the poorest is more than 10% above the metropolitan average. *Nord, Pas-de-Calais, Somme, Aisne* or *Ardennes* are part of it. Conversely, the poorest are underrepresented in *Paris, Hauts-de-Seine* and *Yvelines*, in the same way as 1966, around 20-30%. The Swiss border emerges too – particularly *Haute-Savoie* – in which the poorest tax households are gradually disappearing : the index is 79%, weaker than Paris one.

Figure 6.4.4 maps distortion indices for tax units belonging to the first decile. Section 6.6.2 of the Appendix (Table 6.7) provides these distortion indices for the first and the last 20 *départements* of the panel (ranked according to their values in 2014). The first map reveals the 1966 values. Mirroring the situation of the poorest, one can see that the richest were overrepresented in *Ile-de-France*, and more particularly in *Paris, Hauts-de-Seine* and *Yvelines*. Values were particularly high here : the

Figure 6.4.3 : Distorsion indices for the poorest 50% tax units, 1966 and 2014



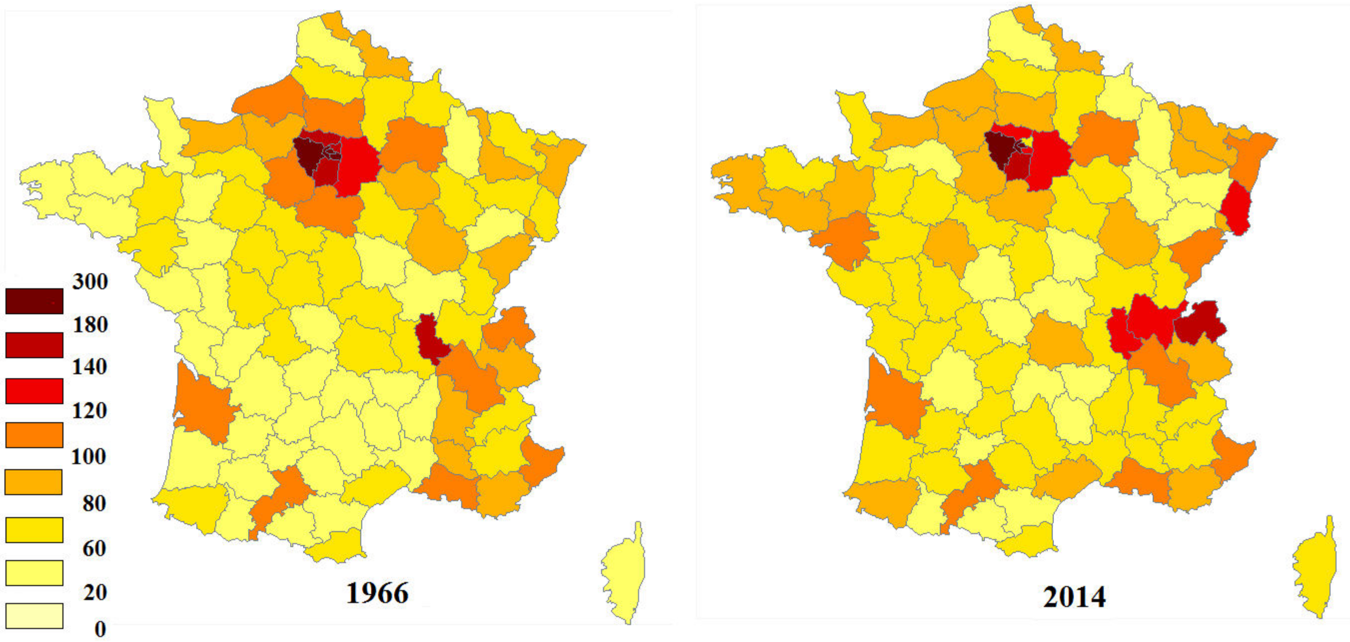
Notes : Distorsion index for decile x is the quotient of the population's share of decile x in one *département* and the same share at the national level. Computations made with the no-capital gains specification. Sample includes 95 *départements* in 1966 and 2014.

proportion of tax units belonging to the first decile was 2.6 times higher in *Hauts-de-Seine* than it was at the metropolitan level (2 times higher in Paris). On the other hand, these were underrepresented in *départements* like *Haute-Loire* or *Creuse*, where indices were close to 30%. One can see that urban *départements* (*Rhône, Gironde, Haute-Garonne, Seine-Lower, Marne*) and the Eastern *départements* were generally close to the metropolitan average.

In 2014, the Paris region is no longer homogeneous. Distortion indices are still high in Paris and the western suburbs, around 200%, but the richest have fled massively *Seine-Saint-Denis*. Globally, the geographical areas are not as marked as before : the South-West is no longer a place where the richest tax units are missing. To sum up, their spatial distribution can be explained by two statements. The first is urbanization issue : *Rhone, Haute-Garonne, Bouches-du-Rhone, Loire-Atlantique, Gironde, Puy-de-Dôme* or *Marne*, in which there are the largest cities outside Paris, emerged on the map as indices are close to the metropolitan average. The second is linked to the Swiss border. *Ain, Haute-Savoie, Bas-Rhin* or *Doubs*, have a large number of tax units belonging to the first decile. The index reached 179% in *Haute-Savoie* (4th higher value), 126% in *Ain* and *Bas-Rhin* (7th and 8th values).

In addition to the cartographic analysis, one can track all distortion indices for a specific *département*. This helps to better understand the transformations of income distributions. We have seen on the previous maps that the situations changed for four specific areas : *Ile-de-France* in which *Seine-Saint-Denis* stands out as a poorer and poorer *département*, South-West which converges towards the metropolitan average but is always late, *Bretagne* in fast development, and the Swiss border which seems to take advantage of the proximity of its rich neighbor. Figure 6.4.5 pictures the distortion indices for four *départements* representative of these geographical areas, namely *Seine-Saint-Denis, Dordogne, Ille-et-Vilaine* and *Haut-Rhin*.

Figure 6.4.4 : Distorsion index for the richest 10% tax units, 1966 and 2014



Notes : Distorsion index for decile x is the quotient of the population's share of decile x in one *département* and the same share at the national level. Computations made with the no-capital gains specification. Sample includes 95 *départements* in 1966 and 2014.

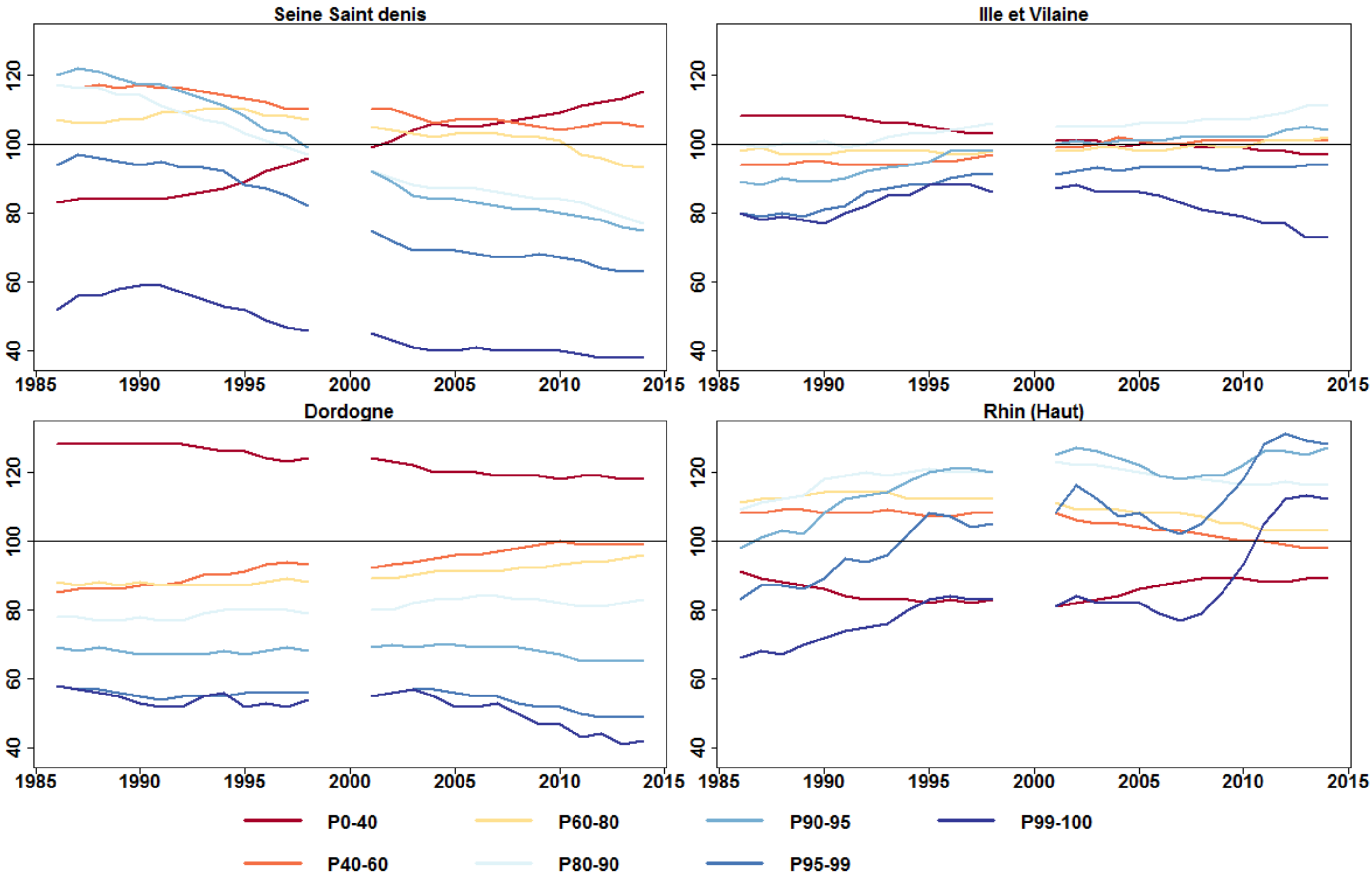
The situation in *Seine-Saint-Denis* is characterized by a very pronounced decline of tax units belonging to the upper middle classes (P80–95). Their distortion indices dropped from 120% in 1986 to 80% in 2014. These tax units were replaced by those belonging to the last four deciles (P0–40), whose distortion indices have gained 40 points in 30 years. One can note that the richest tax units (P99–100) were largely underrepresented in 1986 but are disappearing in this *département*.

Dordogne, in the rural South-West, is characterized by a poor population according to metropolitan standards. The poorest tax units (P0–40) are largely overrepresented but they are replaced gradually by intermediate tax units (P40–80). The share of the upper-middle class (P80–95) remains stable, while the share of the upper-class (P95–100) is lower. Thus, if the South-West is following a catching up process, this is due to the replacement of the poorest by a little less poor tax units.

Ille-et-Vilaine has distorsion indices more homogeneous than the other two *départements*, between 80% and 120%. This *département* is therefore fairly representative of France. However, one can see some changes during these 30 years. The most important is the replacement of the poorest tax units by those of the upper-middle class (P80–99). The distortion index of the former loses 13 points when the one of the latter increases by 15 points. Today, tax units belonging to the second decile are the most overrepresented in this *département*. One can note that the top 1% tax units have been gradually disappearing since 2001.

The situation in *Haut-Rhin* has changed since 1986. This process can be divided into two sub-periods. Between 1986 and 1998, tax units belonging to the last eight deciles were gradually replaced by those belonging to the first two. This trend is particularly clear for the poorest, whose distortion index decreased by 12 points. From 2001 onwards, the share of tax units belonging to the first decile is increasing compared to the share of those belonging to the second. One can also note the

Figure 6.4.5 : Distorsion indices in specific départements, 1986–2014



Notes : Ditorsion index for decile x is the quotient of the population's share of decile x in one *département* and the same share at the national level. Computations made with the no-capital gains specification. Sample includes 95 *départements*.

trend reversal for the poorest, whose population share is starting to increase. *Haut-Rhin* is therefore characterized by an increasingly polarized situation : the proportion of the middle-class is declining while those of the very rich and very poor are on the rise

6.5 Conclusion

In this paper, we have presented a new database concerning distributions of income in the French metropolitan *départements* for the periods 1960–1969 and 1986–2014. This work is based on distributions of income established at the national level for France by Garbinti et al. (2016) as well as the use of unique administrative records collected in the archives of the Finance Ministry. These documents provide, for each year and each *département*, the number of tax units and the total of taxable income for different taxable income brackets. From a technical point of view, we have used the *gpinter* package developed by Blanchet et al. (2017), which aims at estimate the entire distributions. Finally, we use the French departmental demographic database for the 20th century built by Bonnet (2018) to estimate the distributions of income between 1960 and 1969.

We have used two methods according to the two sub-periods 1960–1969 and 1986–2014. Between

1960 and 1969, computations have been more complex insofar as the administrative documents only concern taxable units. To overcome this issue, we have first estimated the total number of tax units for each *département* and each year by using demographic statistics (population by age and married couples) and an econometric model tested on the period 1986–2014. Second, we have estimated the share of income held by the taxable units for each *département* and each year by using the Lorenz curves computed at the national level by Garbinti et al. (2016) as well as the Lorenz curves established at the departmental level for the period 1986–2014. For these two periods, the departmental distributions of income were estimated according to two specifications : one including capital gains, and one not. This follows Garbinti et al. (2016)'s guidelines.

These departmental distributions of income follow the path initiated by Sommeiller et Prize (2014), who were the first to build distributions of income at the subnational level. They shed light on the debate over the "territorial divide", particularly accurate in France. On this point, we have shown that inequalities according to average income per adult have changed over the last 40 years. In detail, the situation seems more homogenized globally, but the spatial segregation observed nowadays in *Ile-de-France* did not exist in the mid-1960s. Moreover, we have brought some ideas on the question of the spatial distribution of the richest or the poorest tax units, by revealing the gentrification of some *département*s. This is the case of the Swiss border, and more particularly *Haute-Savoie*, which hosts more and more top income tax units. Conversely, the poorest are still overrepresented in the South-West : their proportion is 20% higher than the metropolitan average in *Aude, Cantal, Ariège* or *Creuse.*

Although this paper gives a lot of information on intra-departmental inequalities since 1960, the data used are still incomplete. In particular, documents according to the years 1970–1985 and 1915–1959 are not available, even if the income tax existed over these periods. Therefore, the extension of the database is on our agenda. In the case these administrative documents have definitely disappeared, it would be possible to estimate average fiscal incomes per adult using some aggregated statistics collected at the departmental level. With these data, one should analyze the evolution of inter-departmental inequalities over the last 100 years. In addition, an in-depth analysis of intra-departmental inequalities should be conducted. It would reveal the evolution of inequalities for each fractile, giving insights about the spatial distribution of the richest and the poorest, and the evolution of the profile for more than 50 years.

6.6 Appendices

6.6.1 Computations of Departmental Fiscal Incomes Using Regional Accounting

A natural framework for the departmental fiscal income computations between 1960 and 1969 is based on regional accounting, in the same way as the national accounting used by Piketty (2001). The regional accounting exists since 1962 for the 21 former French regions (*Corse* and *Provence-Alpes-Côte d'Azur* counting as one). The raw data are available in Lori (1972) and Lori (1973). Considering that the regional accounts established by INSEE between 1962 and 1969 were reliable, we used these data to compute fiscal incomes between 1960 and 1969 at the departemental level, and deduct the fiscal income of tax-free units.

In the work of Piketty (2001) and Garbinti et al. (2016), the Gross Primary Income of households is central. This includes income from work before redistribution (gross income, employer contributions and worker contributions), gross operating surplus (including that of individual entrepreneurs) and interest and dividends earned by households. Authors used it to deduce the total of fiscal income. Data at the regional level do not provide this aggregate, and items available in the regional accounting do not provide this directly for the 1962–1969 years : labor income is net wages, without social security contributions. To overcome this problem, we used the national Gross Primary Income available in Garbinti et al. (2016). We ensure that the sum of regional Gross Primary Income equals national values by a uniform adjustment. These adjustments were around 20%. Although they appear to be important at first glance, social-contribution rates are almost a linear function of wage income between 1962 and 1969. This suggests that the uniform adjustment made was not far from the reality. Table 6.2 presents the regional distribution of Gross Primary Income obtained for the years 1962–1969.

Three main problems remained : firstly, there was no estimates of Regional Gross Primary Income in 1960 and 1961. We assumed that the ratio of the Gross Primary Income per adult in a region to the Gross Primary Income per adult at the national level was constant between 1960 and 1962. Unsig the ratios in 1962, we computed Gross Primary Income per adult in 1960 and 1961, and then Gross Primary Income. Second, we made the transition from Regional Gross Primary Income to Regional Fiscal Income. To do so we used the national ratio $^{\text{IncomeTax}}/_{\text{GrossPrimaryIncome}}$ for each year. Thus we assumed that this ratio was similar for all regions in each year. Third we allocated Regional Fiscal Income between *départements*. We deducted the fiscal income of tax-free units in each region since we know the fiscal income of taxable units, and assumed that each tax-free unit belonging to a region had the same income.

Nevertheless, estimates were unreliable. This is largely explained by inconsistencies of regional accounting at that time. For example, we obtain in 1966 a negative fiscal income of tax free units in *Ile-de-France*. Table 6.2 shows these inaccuracies : one can note that the percentage of Gross Primary Income for *Ile-de-France* dropped by one point between 1964 and 1966, before rising by 0.5 point in 1968.

Table 6.2 : Regional distribution of gross primary income, 1962–1969

Region	1962	1963	1964	1965	1966	1967	1968	1969
Ile-de-France	28.1%	28.0%	27.9%	27.5%	27.0%	26.9%	27.6%	27.5%
Champagne-Ardennes	2.6%	2.6%	2.6%	2.6%	2.5%	2.5%	2.5%	2.5%
Picardie	3.1%	3.0%	2.9%	2.9%	2.9%	2.9%	2.9%	2.9%
Normandie (Haute)	2.8%	2.9%	2.9%	2.9%	3.0%	3.0%	2.9%	3.0%
Centre	3.5%	3.6%	3.6%	3.8%	3.8%	3.7%	3.7%	3.8%
Normandie (Basse)	2.1%	2.1%	2.0%	2.0%	2.1%	2.2%	2.1%	2.1%
Bourgogne	2.6%	2.6%	2.6%	2.6%	2.6%	2.7%	2.6%	2.7%
Nord-Pas de Calais	6.7%	6.7%	6.6%	6.5%	6.5%	6.5%	6.4%	6.4%
Lorraine	4.5%	4.4%	4.2%	4.1%	4.1%	4.0%	3.9%	3.9%
Alsace	2.6%	2.6%	2.6%	2.6%	2.6%	2.6%	2.6%	2.6%
Franche-Comté	1.8%	1.8%	1.8%	1.7%	1.7%	1.9%	1.8%	1.8%
Pays de la Loire	4.1%	4.1%	4.1%	4.3%	4.3%	4.4%	4.3%	4.3%
Bretagne	4.0%	4.0%	3.9%	3.9%	4.2%	4.2%	4.0%	4.0%
Poitou-Charentes	2.3%	2.4%	2.5%	2.6%	2.4%	2.5%	2.5%	2.5%
Aquitaine	4.1%	4.1%	4.1%	4.2%	4.3%	4.3%	4.5%	4.4%
Midi-Pyrénées	3.4%	3.5%	3.6%	3.7%	3.7%	3.6%	3.6%	3.5%
Limousin	1.3%	1.3%	1.2%	1.3%	1.2%	1.3%	1.2%	1.2%
Rhone-Alpes	8.8%	8.9%	9.0%	9.1%	9.1%	9.1%	8.8%	9.0%
Auvergne	2.3%	2.3%	2.2%	2.2%	2.2%	2.3%	2.3%	2.2%
Languedoc-Roussillon	2.9%	2.7%	2.8%	2.8%	2.8%	2.8%	2.9%	2.9%
Provence-Alpes-Côte-d'Azur + Corse	6.4%	6.5%	6.7%	6.8%	7.0%	6.9%	7.0%	6.7%

Notes : Sample includes 21 regions (*Corse* grouped with *Provence-Alpes-Côte-d'Azur*). Gross Primary Income is the sum of income from work before redistribution, gross operating surplus, interests and dividends earned by households and adjusted by national Gross Primary Income (Garbinti et al., 2016).

6.6.2 Supplementary Materials

Table 6.3 : Annual brackets used in fiscal tabulations

1960	1961–1963	1964	1965	1966	1967–1969	1986–1998	2001–2002	2003–2005	2006–2009	2010–2014
0	0	0	0	0	0	0	0	0	0	0
1,400	1,400	2,400	2,500	2,500	2 ,500	40,000	9,000	7,500	9,400	10,000
2,300	2,500	4,800	5,000	5,000	5,000	50,000	12,000	9,000	11,250	12,000
3,750	4,000	6,500	10,000	7,500	7,500	60,000	19,000	10,500	13,150	15,000
6,500	6,500	10,000	15,000	10,000	10,000	70,000	31,000	12,000	15,000	20,000
9,750	10,000	15,000	20,000	15,000	15,000	80,000	78,000	13,500	16,900	30,000
16,250	15,000	20,000	35,000	20,000	20,000	90,000		15,000	18,750	50,000
32,000	20,000	35,000	50,000	35,000	25,000	100,000		19,000	23,750	100,000
64,000	30,000	45,000	70,000	50,000	30,000	125,000		23,000	28,750	
100,000	60,000	70,000	100,000	70,000	50,000	150,000		31,000	38,750	
150,000	100,000	100,000	200,000	100,000	70,000	200,000		39,000	48,750	
200,000	200,000	200,000	300,000	200,000	100,000	250,000		78,000	97,500	
300,000	300,000	300,000	500,000	300,000	200,000	500,000				
	500,000	500,000		500,000	400,000					

Notes : All thresholds expressed in thousands of Francs until 1998, in Euros from 2001 onwards. In 1963, the eighth threshold is equal to 36,000 instead of 30,000 Francs.

Table 6.4 : Relative fiscal income per adult

Ranking	*Département*	1966	1986	1996	2006	2014
1	Paris	168	155	156	162	162
2	Hauts de Seine	161	147	149	151	152
3	Yvelines	144	144	143	140	141
4	Savoie (Haute)	100	97	97	116	129
5	Essonne	129	132	129	120	116
6	Val de Marne	139	125	126	119	114
7	Alpes Maritimes	107	108	107	112	112
8	Seine et Marne	113	122	118	112	109
9	Ain	89	94	95	102	108
10	Rhône	115	110	110	109	108
11	Val d'Oise	127	124	120	110	106
12	Rhin (Haut)	96	102	107	102	105
13	Savoie	101	100	99	103	103
14	Garonne (Haute)	91	99	100	103	102
15	Isère	105	102	101	103	102
16	Var	97	98	97	102	102
17	Loire Inférieure	93	97	97	100	101
18	Gironde	91	97	97	99	100
19	Rhin (Bas)	99	102	106	100	100
20	Doubs	100	93	96	94	99
76	Loire (Haute)	64	72	78	82	84
77	Pyrénées Orientales	77	80	82	86	84
78	Saône (Haute)	80	82	84	85	84
79	Tarn et Garonne	69	72	78	82	84
80	Seine Saint denis	124	110	102	90	84
81	Indre	78	84	84	83	83
82	Orne	85	83	84	84	83
83	Corse	52	62	72	80	82
84	Dordogne	69	77	80	82	82
85	Lot et Garonne	70	75	80	83	82
86	Marne (Haute)	86	85	85	83	82
87	Meuse	82	86	85	83	82
88	Nièvre	81	86	86	84	82
89	Pas de Calais	80	85	82	81	82
90	Ardennes	92	85	84	82	81
91	Cantal	64	67	73	78	81
92	Aude	69	74	76	80	79
93	Ariège	67	75	77	79	78
94	Lozère	59	68	76	76	78
95	Creuse	58	66	72	74	74

Notes : Table reports the upper and the lower 20 values according to relative fiscal income in 2014. Relative fiscal income is fiscal income per adult relative to the metropolitan average. Sample includes 95 *départements*.

Table 6.5 : Top-to-bottom ratios

Ranking	*Département*	1966	1986	1996	2006	2014
1	Paris	3.3	3.33	3.42	3.45	3.65
2	Hauts de Seine	3.05	3.04	3.18	3.26	3.47
3	Bouches du Rhone	3.22	2.93	3.15	3.15	3.16
4	Val de Marne	2.83	2.84	3	3.11	3.16
5	Yvelines	3.35	2.88	2.98	3.04	3.14
6	Hérault	3.61	3.05	3.22	3.12	3.12
7	Alpes Maritimes	3.3	3.06	3.15	3.12	3.11
8	Seine Saint denis	2.71	2.68	2.91	2.96	3.1
9	Savoie (Haute)	3.26	2.92	3.05	2.98	3.09
10	Marne	3.02	3.02	3.04	3.07	3.08
11	Rhône	3.11	2.89	3.03	3.06	3.08
12	Corse	3.68	3.01	3.31	3.19	3.07
13	Nord	3.29	2.84	3.07	3.05	3.06
14	Garonne (Haute)	3.22	2.94	3.12	3.06	3.04
15	Vaucluse	3.37	2.95	3.1	3.02	3.03
16	Ain	3.33	2.78	2.91	2.91	3.01
17	Gard	3.49	2.94	3.1	3	3.01
18	Gironde	3.4	2.9	3.1	3	2.99
19	Pyrénées Orientales	3.45	2.98	3.14	3	2.99
20	Rhin (Haut)	2.96	2.69	2.81	2.83	2.98
76	Maine et Loire	3.32	2.83	2.96	2.85	2.76
77	Nièvre	3.22	2.82	2.89	2.81	2.76
78	Savoie	3.3	2.8	2.89	2.82	2.76
79	Eure et Loir	3.23	2.75	2.82	2.79	2.75
80	Lozère	3.38	3.2	3.16	2.92	2.75
81	Meuse	3.17	2.75	2.93	2.79	2.75
82	Vosges	3.28	2.71	2.84	2.77	2.75
83	Manche	3.38	2.89	2.96	2.83	2.74
84	Jura	3.24	2.76	2.85	2.72	2.73
85	Landes	3.28	2.96	2.95	2.82	2.73
86	Loir et Cher	3.58	2.78	2.86	2.79	2.73
87	Orne	3.36	2.84	2.93	2.81	2.73
88	Saône et Loire	3.17	2.82	2.91	2.8	2.73
89	Indre	3.37	2.86	2.93	2.79	2.7
90	Loire (Haute)	3.2	2.88	3.02	2.83	2.7
91	Sarthe	3.36	2.76	2.84	2.75	2.7
92	Saône (Haute)	3.18	2.73	2.84	2.72	2.69
93	Marne (Haute)	3.31	2.75	2.88	2.77	2.67
94	Mayenne	3.3	2.84	2.92	2.74	2.65
95	Vendée	3.13	2.83	2.88	2.71	2.6

Notes : Table reports the upper and the lower 20 values according to top-to-bottom ratios in 2014. Top-to-bottom ratio is the ratio of the mean fiscal income of the 5% richest tax units over the mean fiscal income of the 95% poorest tax unit. Sample includes 95 *départements*.

Table 6.6 : Distorsion indices for the poorest 50% tax units

Ranking	*Département*	1966	1986	1996	2006	2014
1	Creuse	145	130	128	125	124
2	Aude	133	121	121	119	119
3	Ariège	120	117	119	117	118
4	Cantal	130	128	128	123	118
5	Pyrénées Orientales	109	112	118	116	118
6	Lot et Garonne	112	118	119	117	116
7	Corse	129	113	121	118	115
8	Dordogne	116	120	118	116	115
9	Gers	141	123	122	118	115
10	Lozère	130	129	122	118	115
11	Pas de Calais	117	115	114	116	115
12	Tarn et Garonne	121	119	121	118	115
13	Aisne	106	109	109	112	113
14	Allier	112	116	113	114	113
15	Ardennes	106	112	112	113	113
16	Gard	125	110	113	112	113
17	Indre	116	118	114	114	113
18	Nièvre	105	113	109	111	113
19	Pyrénées (Hautes)	115	112	113	112	113
20	Vaucluse	107	108	112	112	113
76	Oise	93	93	92	95	97
77	Côte d'Or	97	100	97	96	96
78	Eure et Loir	96	96	93	95	95
79	Loire Inférieure	108	101	101	99	95
80	Loiret	95	93	92	93	95
81	Doubs	97	104	98	98	94
82	Rhin (Bas)	100	98	89	92	94
83	Isère	93	95	95	94	93
84	Rhône	73	92	91	93	93
85	Savoie	100	98	97	95	93
86	Ain	116	100	99	93	90
87	Rhin (Haut)	101	94	87	90	90
88	Val d'Oise	72	76	78	84	90
89	Val de Marne	57	79	78	83	87
90	Seine et Marne	88	80	81	83	86
91	Paris	64	81	81	80	82
92	Essonne	72	71	71	76	81
93	Savoie (Haute)	98	98	101	84	79
94	Hauts de Seine	61	74	71	71	73
95	Yvelines	71	67	66	68	71

Notes : Table reports the upper and the lower 20 values according to distortion indices of the 50% poorest tax units in 2014. Distorsion Index for decile x is the quotient of the population's share of decile x in one *département* and the same share at the national level. Sample includes 95 *départements*.

Table 6.7 : Distorsion indices for the richest 10% tax units

Ranking	*Département*	1966	1986	1996	2006	2014
1	Hauts de Seine	263	193	197	205	217
2	Yvelines	204	227	226	221	217
3	Paris	224	181	179	193	198
4	Savoie (Haute)	101	100	93	143	179
5	Essonne	177	195	185	167	154
6	Val de Marne	188	155	150	140	139
7	Ain	69	89	92	112	126
8	Rhin (Haut)	68	89	111	109	126
9	Seine et Marne	133	153	147	134	126
10	Rhône	142	118	121	121	125
11	Val d'Oise	176	165	158	135	125
12	Garonne (Haute)	104	109	111	116	119
13	Alpes Maritimes	103	116	106	111	109
14	Isère	104	99	101	106	107
15	Rhin (Bas)	80	88	107	105	106
16	Marne	111	116	107	103	103
17	Doubs	87	73	83	85	102
18	Loire Inférieure	74	84	91	97	102
19	Bouches du Rhone	106	100	95	98	101
20	Gironde	104	101	97	98	100
76	Lot et Garonne	48	70	64	64	60
77	Ardennes	75	62	63	60	59
78	Meuse	57	58	66	60	59
79	Pas de Calais	54	57	61	60	59
80	Tarn et Garonne	54	68	61	61	59
81	Loire (Haute)	34	58	57	59	57
82	Pyrénées (Hautes)	59	71	67	64	57
83	Vosges	68	59	64	62	57
84	Allier	63	65	64	61	56
85	Dordogne	46	64	62	62	56
86	Saône (Haute)	55	56	60	60	56
87	Aude	47	63	57	58	55
88	Orne	65	63	64	61	55
89	Nièvre	57	63	64	60	54
90	Cantal	42	59	57	57	53
91	Indre	60	61	61	57	52
92	Lozère	40	58	62	59	52
93	Marne (Haute)	64	58	61	57	51
94	Ariège	54	63	59	55	50
95	Creuse	32	54	54	51	43

Notes : Table reports the upper and the lower 20 values according to distortion indices of the 10% richest tax units in 2014. Distorsion Index for decile x is the quotient of the population's share of decile x in one *département* and the same share at the national level. Sample includes 95 *départements*.

Conclusion

Cette thèse n'aurait pas pu se faire sans la création de nouvelles bases de données historiques portant à la fois sur l'économie et la démographie au niveau départemental en France. Ces bases de données uniques ont permis de mettre en évidence des résultats originaux, qui ont été présentés dans les chapitres 1 à 4. Cependant, elle n'est qu'une étape dans le projet de recherche qui se dessine, en l'occurence la description et l'analyse des inégalités spatiales en France sur longue période. Dans cette conclusion, je rappellerai donc les principaux résultats déjà présentés, ainsi que les limites actuelles de ces travaux. Enfin, j'esquisserai les pistes de recherche qui sont aujourd'hui en voie d'exploration.

Des résultats nouveaux

Dans le Chapitre 1, j'ai analysé l'évolution de la distribution spatiale de la population française, et ce sur la période 1851–2014. J'ai montré dans un premier temps, en analysant le simple indice de Gini, que la population était plus inégalitairement répartie qu'en 1851. Puis, en utilisant un panel plus large d'indicateurs d'inégalités, j'ai décrit ce processus comme une succession de trois phases. Si dans les deux premières l'augmentation de la part de la population résidant dans les régions les plus densément peuplées a contribué à augmenter les inégalités, ce n'est pas le cas de la troisième. En effet, entre 1968 et 2014, les régions un peu moins densément peuplées ont vu leur part dans la population totale augmenter, au détriment à la fois des régions très denses et des régions peu denses. Ce résultat vient démontrer l'importance du choix des indicateurs d'inégalités : un indicateur agrégé comme l'indice de Gini ne permet pas de cerner parfaitement les phénomènes à l'oeuvre. Dans un deuxième temps, j'ai analysé la dynamique de la répartition spatiale de la population selon l'âge. J'ai ainsi montré que le profil par âge des indicateurs d'inégalités avait largement changé depuis 1851. A cette époque, les plus âgés étaient plus inégalitairement répartis que les populations de moins de 60 ans, phénomène s'expliquant par des différences départementales de mortalité. Aujourd'hui, ce sont les adultes âgés de 30 à 40 ans qui sont les plus inégalitairement répartis, alors que les plus âgés sont plus égalitairement répartis sur le territoire national car peu contraints par la localisation des emplois. Enfin, ce chapitre m'a permis de montrer que les structures par âge des populations départementales sont de plus en plus différentes, et ce depuis la fin de la seconde guerre mondiale. Certains départements du Sud-Ouest ont une part de jeunes actifs bien plus faible que cette même part au niveau national, leur tissu économique reposant essentiellement sur les plus âgés.

Dans le Chapitre 2, co-écrit avec Hippolyte d'Albis, nous nous sommes intéressés aux inégalités interdépartementales de mortalité sur la période 1806–2014, en utilisant à la fois la nouvelle base de données portant sur la démographie départementale, mais aussi les données calculées par Bonneuil (1997) pour le 19$^{\text{ème}}$ siècle. Pour cela, nous avons procédé à une description des inégalités d'espérance de vie à différents âges, en utilisant là encore un large panel d'indicateurs. Cela nous a permis d'identifier une période, 1880–1980, durant laquelle les inégalités de mortalité ont diminué en parallèle d'une augmentation continue de l'espérance de vie au niveau national. Durant cette période, seules les deux guerres mondiales ont momentanément interrompu le processus, sans le remettre en cause. Entre 1806 et 1880, la France a connu une succession de hausses et de baisses de l'espérance de vie, ainsi qu'une succession de hausses et de baisses des inégalités de mortalité. Entre 1980 et 2014, les inégalités de mortalité ont recommencé à augmenter, l'espérance de vie augmentant plus vite là où elle était déjà plus élevée que la moyenne nationale. Au-delà de cette analyse générale des inégalités spatiales, nous avons identifié les zones dans lesquelles l'espérance de vie était plus faible ou plus élevée que la moyenne nationale, et ce pour différentes sous-périodes. Ainsi, il s'avère que la Bretagne a de tout temps été une zone où la mortalité était plus élevée. Au contraire, le Nord et le Nord-Est de la France, ainsi que le Sud-Est, ont vu leur position relative changer : les premiers ont vu leur position relative se dégrader fortement, alors que le Sud-Est est aujourd'hui une région où l'espérance de vie est supérieure à la moyenne nationale. Ce résultat est vrai également pour les départements urbains, qui souffraient d'une "pénalité urbaine" au 19$^{\text{ème}}$ siècle du fait d'épidémies plus fréquentes et plus dévastatrices.

Dans le Chapitre 3, j'ai poursuivi mon analyse des inégalités spatiales en procédant à une description des inégalités de revenu et de bien-être sur la période 1922–2014. Les données dont je disposais grâce aux deux bases de données départementales ont été complétées, afin de disposer à la fois de la population et du revenu total de chaque département sur cette période. Cela m'a permis de calculer le revenu par habitant de chacun d'entre eux. Dans un premier temps, j'ai montré que les inégalités de revenu ont continuellement baissé depuis le début des années 1920. Cette baisse peut être encore une fois séquencée en 3 étapes. Les inégalités ont diminué très lentement jusqu'en 1950, puis à un rythme bien plus soutenu jusqu'en 1980. De 1980 à nos jours, les inégalités ont à nouveau diminué à un rythme très lent. La période 1950-1980 est particulièrement intéressante : la baisse des inégalités de revenu par habitant s'est faite en parallèle d'une répartition plus homogène du revenu total sur le territoire national. J'ai donc qualifié cette phase de "30 glorieuses des inégalités spatiales". Dans ce chapitre, j'ai également calculé les "revenus ajustés de la mortalité" en utilisant la méthodologie proposée par Fleurbaey et Gaulier (2009). Ces revenus incorporent une monétarisation des différences de mortalité entre les départements. En d'autres termes, les départements où la mortalité est élevée subissent une pénalité monétaire. Ainsi, j'ai pu montrer qu'en incorporant ces inégalités de mortalité, les inégalités spatiales n'avaient pas baissé entre 1922 et 1950 mais augmenté. Ce résultat s'explique par la forte mortalité des départements urbains, qui étaient également les départements aux revenus les plus élevés. La convergence des conditions de mortalité, très forte sur la période, a donc entraîné une hausse des inégalités de bien-être, avant que celles-ci ne diminuent après 1950. Dans un dernier temps, j'ai montré de quelle manière les départements du Nord et du Nord-Est de la France ont vu leur position relative se dégrader fortement entre 1922 et 2014. Sur la même période, le Sud-Ouest du pays a profité d'un processus de rattrapage important.

Dans le Chapitre 4, je me suis détaché de l'analyse des inégalités spatiales sur longue période pour travailler sur un épisode tragique de l'histoire récente française, celui des flux internes de population durant la seconde guerre mondiale. On connaît assez bien les flux massifs de population qui ont eu lien entre mai et juin 1940 grâce aux travaux des historiens. Cependant, les sources statistiques à ma disposition m'ont permis de renouveler cette analyse et de dresser un portrait exhaustif des flux sur la période 1939–1946. Dans un premier temps, j'ai proposé une méthode permettant de calculer les flux internes annuels de population pour chaque période intercensitaire. En effet, la méthodologie généralement utilisée n'était pas pertinente pour mener à bien ce travail descriptif. Grâce à cette nouvelle méthodologie, basée notamment sur la connaissances des décès par cause au niveau local, j'ai calculé les populations départementales de chaque année et déduit les flux migratoires apparents. J'ai pu ainsi dresser une cartographie de ces flux pour chaque année de guerre. Leur description montre que les réfugiés se sont d'abord orientés vers le Centre-Ouest de la France en 1940, avant de prendre la direction des régions situées au-delà de la ligne de démarcation. Les départements du Nord-Est de la France, puis plus généralement ceux au nord de la ligne de démarcation, ont vu leur population partir progressivement. A partir de 1943, les populations réfugiées sur la côte méditerranéenne ont fui leur terre d'exil provisoire pour se réfugier dans l'intérieur des terres, avant de regagner progressivement leur département d'origine. La seconde guerre mondiale a laissé une trace forte dans la démographie départementale, puisque les réfugiés de certains départements du Nord de la France n'étaient pas encore rentrés en 1946, alors que la population du Sud-Ouest avait augmenté de façon pérenne du fait de ces mouvement migratoires.

Limites des travaux actuels

La principale limite de mes travaux à l'heure actuelle porte sur la connaissance des populations départementales pour chaque année intercensitaire. Dans la construction de la base de données portant sur la démographie départementale, les populations par âge sont connues avec précision à la date de chaque recensement. Avec le temps, la qualité de ces recensements s'est par ailleurs largement améliorée. Néanmoins, ces recensements ont été menés au $20^{\text{ème}}$ siècle et au début $21^{\text{ème}}$ siècle de façon intermittente, ce qui ne permet pas de connaître les populations de chaque année. Pour pallier ce manque, le protocole de la Human Mortality Database propose de calculer les populations au 1^{er} janvier de chaque année en utilisant un certain nombre de méthodes. La principale consiste à répartir les flux migratoires apparents entre chaque recensement de manière proportionnelle au temps écoulé. Cette méthode peut être appropriée au niveau national, mais l'est moins au niveau local dans la mesure où les flux migratoires sont plus importants. Ce problème a été démontré dans le Chapitre 4, portant sur l'exode durant la seconde guerre mondiale. Dans ce chapitre, une autre méthode a été utilisée pour calculer les populations départementales annuelles, ce qui m'a permis de disposer d'estimations plus précises des flux migratoires. De manière générale, cette question devra être approfondie à l'avenir car les estimations de populations départementales pour chaque année sont largement reprises dans les travaux économiques, notamment l'estimation des revenus départementaux de chaque année sur la période 1922–1959 et 1970–1985. D'autres sources statistiques pourront être utilisées pour cela, comme par exemple le nombre de foyers fiscaux qui sont eux disponibles sur une base annuelle.

Au-delà de cette question de l'estimation des populations départementales pour chaque période intercensitaire, la question des migrations interdépartementales n'a pas été étudiée dans cette thèse. A peine a-t-elle été effleurée dans le Chapitre 1, pour dissocier les évolutions de population provenant du mouvement naturel et du mouvement migratoire. Pourtant, les données disponibles dans la base de données démographique permettrait de procéder à une description des flux migratoires interdépartementaux sur longue période. Actuellement, la seule limitation à ces travaux potentiels est que ces flux migratoires ne sont disponibles qu'à la date de chaque recensement à partir de 1906. Néanmoins, ils recèlent une masse importante d'informations permettant d'expliquer à la fois le sens, l'intensité, les différences entre les hommes et les femmes, et les variations temporelles de ces migrations.

Troisièmement, les travaux concernant les inégalités spatiales présentés dans cette thèse se limitent à l'étude des inégalités interdépartementales. Ainsi, j'ai procédé à l'analyse des inégalités de densités de population dans le Chapitre 1, des inégalités d'espérance de vie dans le Chapitre 2, des inégalités de revenus par habitant (ajustés, ou non, par la mortalité) dans le Chapitre 3. Néanmoins, les bases de données construites me permettront par la suite de procéder à une analyse plus fine des inégalités, en m'intéressant aux inégalités intradépartementales. La base de données portant sur la démographie départementale permet de disposer de la distribution des âges de décès dans chaque département, et ce pour chaque année et chaque sexe. Ainsi, les indicateurs d'inégalités calculés pour la distribution départementale des espérances de vie pourront être calculés dans chaque département, afin de comprendre dans quelle mesure la baisse des inégalités totales provient d'une baisse des inégalités interdépartementales ou d'une baisse des inégalités intradépartementales. La base de données portant sur la distribution des revenus dans chaque département pour la période 1960–1969 et 1986–2014 ouvre également la voie à ce type de travaux. En effet, il est possible de calculer les indicateurs d'inégalités dans chacun des départements et pour chacune des années. Cela permettra de comprendre en quoi l'évolution récente des inégalités au niveau national provient de l'évolution des inégalités interdépartementales ou intradépartementales. Par ailleurs, cette base de données permettra de savoir où se situent les foyers fiscaux définis comme les plus pauvres où les plus riches au niveau national, et d'analyser la dynamique de leur répartition spatiale.

Enfin, les trois premiers chapitres de cette thèse sont essentiellement descriptifs. Après avoir construit les bases de données portant sur la démographie et les revenus départementaux, le but était de dresser un portrait le plus clair possible de l'évolution des inégalités spatiales, tant sur le plan de la distribution de la population sur le territoire national qu'en ce qui concerne la mortalité ou encore les revenus par habitant. En effet, ces travaux n'avaient encore jamais été entrepris sur une période aussi longue, ce qui rendait ce travail descriptif indispensable. Pour autant, il est essentiel aujourd'hui de comprendre les raisons de ces évolutions, que ce soit la distribution plus inégale de la population, la baisse des inégalités de mortalité ou la baisse des inégalités de revenus par habitant. Pour procéder à ces travaux analytiques, il sera intéressant de coupler l'intégralité des variables obtenues pour voir en quoi elles interagissent les unes avec les autres. Parmi les nombreuses questions qui se posent, on retiendra par exemple les deux suivantes. La baisse des inégalités de revenus par habitant a-t-elle permis la baisse des inégalités de mortalité ? Les flux migratoires de population peuvent-ils s'expliquer par les différences interdépartementales de revenus ?

Perspectives futures de recherche

Dans la section précédente, j'ai esquissé les projets qui pourront être menés en utilisant les bases de données déjà construites. Néanmoins, mes travaux en archives m'ont permis de mettre à jour un grand nombre de données brutes qui n'ont à ce jour pas été exploitées. Ces données feront l'objet d'études dans les mois et les années qui viennent. Pour conclure cette thèse, je voudrais présenter trois d'entre elles, à savoir les données de fécondité, les données concernant les causes de décès et les données de richesse.

La base de données portant sur la démographie départementale à été construite à partir de statistiques brutes concernant l'âge de décès des défunts de chaque département et chaque sexe, et ce pour chaque année. C'est une des deux composantes de ce que l'on appelle le mouvement naturel. Dans les publications où ces données ont été collectées se trouvent également les statistiques brutes portant sur le nombre de naisances selon l'âge de la mère, et ce pour chaque département et chaque année entre 1906 et 2014. Identifiées, ces statistiques vont être collectées puis digitalisées afin de pouvoir construire les tables de fertilité, dans la lignée des tables de mortalité déjà construites. Mises ensemble, elles permettront de dresser le portrait de la fin de la transition démographique au niveau local en France, mais aussi de comprendre les ressorts des différences de fécondité selon les différentes régions. Un exemple particulièrement connu des démographes français est l'apparition d'un "croissant fertile" dans le Nord à la suite de la seconde guerre mondiale. Les pistes scientifiques pour l'expliquer sont à ce jour encore assez floues.

De la même manière, les publications dans lesquelles les données de mortalité selon l'âge ont été collectées fournissent les données de mortalité selon la cause de décès, et ce pour chaque département depuis 1906, et selon chaque sexe depuis les années 1920. Ces données ont à ce jour été digitalisées mais leur utilisation invite à la plus grande prudence. En effet, à la différence des décès selon l'âge, les décès selon la cause sont classés selon des nomenclatures qui ont largement changé sur toute la période d'étude. Une analyse sur longue période nécessite donc dans un premier temps le reclassement de ces décès dans une nomenclature unique pouvant être suivie dans le temps. Pour cela, les travaux de Vallin et Meslé (1988) seront particulièrement utiles puisque ces auteurs ont procédé à cette reclassification au niveau national. Par ailleurs, un autre écueil concerne la part des décès de cause inconnue ou mal spécifiée, notamment durant la première moitié du 20$^{\text{ème}}$ siècle. Ces décès ne peuvent pas être supprimés ni répartis proportionnellement, ce qui oblige à trouver une méthodologie permettant de procéder à leur répartition. Malgré ces potentiels problèmes, ces données permettront à terme une analyse riche de la convergence des conditions de mortalité au niveau local, en identifiant les causes de décès qui ont eu le plus d'importance dans ce processus. Par ailleurs, dans une approche plus actuelle, elles permettront aussi de comprendre pourquoi les inégalités spatiales de mortalité sont reparties à la hausse depuis les années 1980.

Enfin, les archives du ministère de l'Economie et des Finances contiennent un grand nombre de données fiscales au niveau local qui n'ont à ce jour pas encore été utilisées. Les données concernant l'impôt sur le revenu ont été récupérées pour calculer à la fois les distributions départementales de revenu pour les périodes 1960–1969 et 1986–2014, et les revenus départementaux pour les périodes 1922–1959 et 1970–1985. Néanmoins, ce ne sont pas les seules disponibles. Des données départementales concernant les déclarations de succession sur la période 1900–1960 existent également : ces statistiques fournissent, pour chaque département et quasiment chaque année, le nombre de successions et l'actif total selon différentes tranches d'actif successoral net. Elles fournissent également, pour chaque département et pour quelques années, le nombre de successions et le montant de l'actif successoral total selon l'âge du défunt. Elles fournisent enfin, pour chaque département et pour quelques années, l'actif successoral total selon le type d'actif (actions, obligations, biens immobiliers, or...). Ces données, grâce à leur degré de détail, doivent permettre de reconstituer la richesse de chaque département sur la période 1900–1960. Ces travaux s'inscriront alors dans la lignée des travaux de Garbinti et al. (2017), qui ont reconstitué cette richesse total au niveau national sur très longue période. Grâce à cette nouvelle base de données, il sera plus facile de comprendre les différences spatiales de revenus par habitant entre 1922 et 1960, mais aussi de procéder à une analyse historique de l'évolution des inégalités spatiales de richesse par habitant.

Bibliography

Alary, E. (2013). *L'exode*. Perrin.

Alary, E., B. Vergez-Chaignon, and G. Gauvin (2006). *Les Français au quotidien, 1939-1949*. Perrin.

Alvaredo, F. (2009). Top Incomes and Earnings in Portugal 1936–2005. *Explorations in Economic History 46*(4), 404–417.

Alvaredo, F., A. Atkinson, L. Chancel, T. Piketty, E. Saez, G. Zucman, et al. (2016). Distributional National Accounts (DINA) Guidelines: Concepts and Methods used in WID.World. *WID.World Working Paper 2*.

Alvaredo, F. and E. Saez (2009). Income and Wealth Concentration in Spain from a Historical and Fiscal Perspective. *Journal of the European Economic Association 7*(5), 1140–1167.

Ariès, P. (1948). *Histoire des populations françaises et de leurs attitudes devant la vie depuis le XVIIIe siècle*. Éditions Self.

Atkinson, A. B. (2005). Top Incomes in the UK over the 20th Century. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 168*(2), 325–343.

Atkinson, A. B. and W. Salverda (2005). Top Incomes in the Netherlands and the United Kingdom over the 20th Century. *Journal of the European Economic Association 3*(4), 883–913.

Ayuda, M. I., F. Collantes, and V. Pinilla (2010). Long-run Regional Population Disparities in Europe During Modern Economic Growth: a Case Study of Spain. *The Annals of Regional Science 44*(2), 273–295.

Baccaïni, B. (1999). Analyse des migrations internes et estimation du solde migratoire externe au niveau local à l'aide des données censitaires. *Population (French Edition) 54*(4/5), 801–815.

Baccaïni, B. (2001). Les migrations internes en France de 1990 à 1999 : l'appel de l'Ouest. *Économie et statistique 344*(1), 39–79.

Baccaïni, B., D. Courgeau, and G. Desplanques (1993). Les migrations internes en France de 1982 à 1990. Comparaison avec les périodes antérieures. *Population (French Edition) 48*(6), 1771–1789.

Baccaïni, B. and C. Dutreuilh (2007). Inter-Regional Migration Flows in France over the Last Fifty Years. *Population (English Edition)*, 139–155.

Baccaïni, B. and D. Levy (2009). Recensement de la population de 2006. Les migrations entre départements: le Sud et l'Ouest toujours très attractifs. *Insee première* (1248), 4.

Badía-Miró, M., J. Guilera, and P. Lains (2012). Regional Incomes in Portugal: Industrialisation, Integration and Inequality, 1890-1980. *Revista de Historia Económica-Journal of Iberian and Latin American Economic History 30*(2), 225–244.

Barbieri, M. (2013). La mortalité départementale en France. *Population 68*(3), 433–479.

Barbieri, M. and N. Ouellette (2012). La démographie du Canada et des Etats-Unis des années 1980 aux années 2000. *Population 67*(2), 221–328.

Barro, R. J. and J. W. Friedman (1977). On Uncertain Lifetimes. *Journal of Political Economy 85*(4), 843–849.

Barro, R. J. and X. Sala-i Martin (1992). Convergence. *Journal of Political Economy 100*(2), 223–251.

Bazot, G. (2014). Interregional Inequalities, Convergence, and Growth in France from 1840 to 1911. *Annals of Economics and Statistics/Annales d'Économie et de Statistique* (113/114), 309–345.

Becker, G. S., T. J. Philipson, and R. R. Soares (2005). The Quantity and Quality of Life and the Evolution of World Inequality. *American Economic Review 95*(1), 277 – 291.

Beeson, P. E., D. N. DeJong, and W. Troesken (2001). Population Growth in US Counties, 1840–1990. *Regional Science and Urban Economics 31*(6), 669–699.

Berry, C. R. and E. L. Glaeser (2005). The Divergence of Human Capital Levels across Cities. *Papers in Regional Science 84*(3), 407–444.

Beyers, W. B. and P. B. Nelson (2000). Contemporary Development Forces in the Nonmetropolitan West: New Insights from Rapidly Growing Communities. *Journal of Rural Studies 16*(4), 459–474.

Blanc, M., B. Schmitt, and E. Ambiaud (2007). Orientation économique et croissance locale de l'emploi dans les bassins de vie des bourgs et petites villes. *Economie et statistique 402*(1), 57–74.

Blanchet, T., J. Fournier, and T. Piketty (2017). Generalized Pareto Curves: Theory and Applications. *WID.World Working Paper*.

Bolt, J., R. Inklaar, H. de Jong, and J. L. van Zanden (2018). Rebasing 'Maddison': New Income Comparisons and the Shape of Long-run Economic Development. *GGDC Research Memorandum 174*.

Bonnet, F. (2018a). Beyond the Exodus of May-June 1940: Internal Migrations in France during the Second World War. *mimeo*.

Bonnet, F. (2018b). Computations of French Lifetables by Département, 1901–2014. *mimeo*.

Bonnet, F. (2018c). Spatial Distribution of Populations by Age, 1851–2014. *mimeo*.

Bonnet, F. and A. Sotura (2018). Computations of French Distributions of Income by Département, 1960–2014. *mimeo*.

Bonneuil, N. (1997). *Transformation of the French Demographic Landscape 1806-1906.* Oxford England Clarendon Press.

Bontje, M., S. Musterd, et al. (2012). *Understanding Shrinkage in European Regions.* Alexandrine Press.

Bourdelais, P. (1993). *L'Age de la vieillesse.* Odile Jacob.

Bourdieu, J., L. Kesztenbaum, and G. Postel-Vinay (2014). *L'enquete TRA, histoire d'un outil, outil pour l'histoire: Tome I. 1793–1902.* INED.

Breton, D., M. Barbieri, H. d'Albis, and M. Mazuy (2017). L'évolution démographique récente de la france: de forts contrastes départementaux. *Population 72*(4), 583–651.

Brown, D. and P. Rees (2006). Trends in Local and Small Area Mortality and Morbidity in Yorkshire and the Humber: Monitoring Health Inequalities. *Regional Studies 40*(5), 437–458.

Brunet, G. and A. Bideau (2000). Démographie historique et généalogie. *Annales de Démographie Historique 14*(2), 9–38.

Buyst, E. (2010). Reversal of Fortune in a Small, Open Economy: Regional GDP in Belgium, 1896–2000. *Rivista di storia economica 26*(1), 75–92.

Caruana-Galizia, P. (2013). Estimating French Regional Income: Departmental per Capita Gross Value Added, 1872–1911. In *Research in Economic History*, pp. 71–95. Emerald Group Publishing Limited.

Charpentier, A. and E. Gallic (2018). Étude de la démographie française du XIXe siècle à partir de données collaboratives de généalogie. *https://hal.archives-ouvertes.fr/hal-01724269/*.

Cheung, S. L. K., J.-M. Robine, E. J.-C. Tu, and G. Caselli (2005). Three Dimensions of the Survival Curve: Horizontalization, Verticalization, and Longevity Extension. *Demography 42*(2), 243–258.

Coale, A. J. (2017). *The Decline of Fertility in Europe.* Princeton University Press.

Coale, A. J., P. Demeny, and B. Vaughan (2013). *Regional Model Life Tables and Stable Populations: Studies in Population.* Elsevier.

Combes, P.-P., G. Duranton, and L. Gobillon (2012). The Costs of Agglomeration: Land Prices in French Cities. *CEPR Discussion Paper No. 9240*.

Combes, P.-P., M. Lafourcade, J.-F. Thisse, and J.-C. Toutain (2011). The Rise and Fall of Spatial Inequalities in France: A Long-Run Perspective. *Explorations in Economic History 48*(2), 243–271.

Courgeau, D. (1978). Les migrations internes en France de 1954 à 1975. I Vue d'ensemble. *Population (French Edition) 33*(3), 525–545.

Courgeau, D. and E. Lelièvre (2004). Estimation des migrations internes de la période 1990–1999 et comparaison avec celles des périodes antérieures. *Population (French Edition) 59*(5), 797–804.

Cowell, F. (2011). *Measuring Inequality.* Oxford University Press.

Currie, J., H. Schwandt, and J. Thuilliez (2018). Pauvreté, Egalité, Mortalité: Mortality (in) Equality in France and the United States. Technical report, National Bureau of Economic Research.

Daguet, F. (2006a). Dans quelles régions meurt-on le plus tard au début du XXIème siècle ? *Insee première* (1114).

Daguet, F. (2006b). *Données de démographie régionale de 1954 à 1999.* INSEE.

d'Albis, H. and F. Bonnet (2018a). Inequalities in Life Expectancy and the Global Welfare Convergence. *Economics Letters 168*, 49–51.

d'Albis, H. and F. Bonnet (2018b). Spatial Inequalities of French Life Expectancy, 1806–2014. *mimeo*.

d'Albis, H. and F. Collard (2013). Age Groups and the Measure of Population Aging. *Demographic Research 29*, 617–640.

d'Albis, H., L. J. Esso, and H. Pifarré i Arolas (2014). Persistent Differences in Mortality Patterns across Industrialized Countries. *PloS one 9*(9), e106176.

Davezies, L. (2008). *La France et ses territoires, la circulation invisible des richesses.* Seuil.

Dean, K. G. (1988). Interregional Flows of Economically Active Persons in France, 1975–1982. *Demography 25*(1), 81–98.

Diamond, H. (2008). *Fleeing Hitler: France 1940.* Oxford University Press.

Dupâquier, J. and J.-P. Bardet (1988). *Histoire de la population française*, Volume 4. Presses universitaires de France.

Edwards, R. D. (2012). The Cost of Uncertain Life Span. *Journal of Population Economics 26*(4), 1485–1522.

Edwards, R. D. and S. Tuljapurkar (2005). Inequality in Life Spans and a New Perspective on Mortality Convergence across Industrialized Countries. *Population and Development Review 31*(4), 645–674.

Elo, I. T. and S. H. Preston (1996). Educational Differentials in Mortality: United States, 1979–1985. *Social science & medicine 42*(1), 47–57.

Enflo, K. and J. R. Rosés (2015). Coping With Regional Inequality in Sweden: Structural Change, Migrations, and Policy, 1860–2000. *The Economic History Review 68*(1), 191–217.

Ezzati, M., A. B. Friedman, S. C. Kulkarni, and C. J. Murray (2008). The Reversal of Fortunes: Trends in County Mortality and Cross-County Mortality Disparities in the United States. *PLoS Med 5*(4), e66.

Felice, E. and G. Vecchi (2015). Italy's Growth and Decline, 1861–2011. *Journal of Interdisciplinary History 45*(4), 507–548.

Fletcher, T. W. (1961). The Great Depression of English Agriculture 1873–1896. *The Economic History Review 13*(3), 417–432.

Fleurbaey, M. and G. Gaulier (2009). International Comparisons of Living Standards by Equivalent Incomes. *The Scandinavian Journal of Economics 111*(3), 597–624.

Florida, R. (2002). The Economic Geography of Talent. *Annals of the Association of American geographers 92*(4), 743–755.

Fol, S. (2012). Urban Shrinkage and Socio-Spatial Disparities: Are the Remedies Worse than the Disease? *Built Environment 38*(2), 259–275.

Fourastié, J. (1979). *Les trente glorieuses ou la Révolution invisible de 1946 à 1975.* Fayard.

Francart, G. (1983). Le rééquilibrage démographique de la France. *Economie et statistique 153*(1), 35–46.

Fries, J. F. (2002). Aging, Natural Death, and the Compression of Morbidity. *Bulletin of the World Health Organization 80*(3), 245–250.

Gächter, M. and E. Theurl (2011). Health Status Convergence at the Local Level: Empirical Evidence from Austria. *International Journal for Equity in Health 10*(1), 34.

Galjaard, R., L. Van Wissen, and K. Van Dam (2012). European Regional Population Decline and Policy Responses: Three Case Studies. *Built Environment 38*(2), 293–302.

Garbinti, B., J. Goupille-Lebret, and T. Piketty (2016). Income Inequality in France, 1900-2014: Evidence from Distributional National Accounts (DINA). *WID.World Working Paper*.

Garbinti, B., J. Goupille-Lebret, and T. Piketty (2017). Accounting for Wealth Inequality Dynamics: Methods, Estimates and Simulations for France (1800–2014).

Geary, F. and T. Stark (2002). Examining Ireland's Post-Famine Economic Growth Performance. *The Economic Journal 112*(482), 919–935.

Glaeser, E. L. (1999). Learning in Cities. *Journal of Urban Economics 46*(2), 254–277.

Goreux, L.-M. (1956). Les migrations agricoles en France depuis un siécle et leur relation avec certains facteurs économiques. *Etudes et conjoncture 11*(4), 327–376.

Graham, D. J. (2007). Variable Returns to Agglomeration and the Effect of Road Traffic Congestion. *Journal of Urban Economics 62*(1), 103–120.

Gravier, J.-F. (1947). *Paris et le désert français: décentralisation, équipement, population.* Flammarion.

Haines, M. R. (2001). The Urban Mortality Transition in the United States, 1800–1940. *Annales de Démographie Historique 1*, 33–64.

Houdaille, J. (1971). Les migrations intérieures en France à la fin du XVIIIe et au début du XIXe siècle. *Population 26*(4), 743–744.

Huber, M. (1931). *La population de la France pendant la guerre.* Presses Universitaires de France.

Hunt, J. (1992). The Impact of the 1962 Repatriates from Algeria on the French Labor Market. *ILR Review 45*(3), 556–572.

Illsley, R. and J. Le Grand (1993). Regional Inequalities in Mortality. *Journal of Epidemiology and Community Health 47*(6), 444–449.

Jackson, J. (2004). *The Fall of France: The Nazi Invasion of 1940.* Oxford University Press.

Janssen, F., A. van den Hende, J. A. de Beer, and L. J. van Wissen (2016). Sigma and Beta Convergence in Regional Mortality: A Case Study of the Netherlands. *Demographic Research 35*, 81–116.

Jayet, H., J.-P. Puig, and J.-F. Thisse (1996). Enjeux économiques de l'organisation du territoire. *Revue d'économie politique*, 127–158.

Jones, C. I. and P. J. Klenow (2010). Beyond GDP? Welfare Across Countries and Time. Technical report, National Bureau of Economic Research.

Joseph, K., L. Huang, S. Dzakpasu, and C. Mc Court (2009). Regional Disparities in Infant Mortality in Canada: a Reversal of Egalitarian Trends. *BMC Public Health 9*(1), 4.

Kesztenbaum, L. (2014). L'étude des migrations grace aux registres matricules militaires. *Popolazione e Storia 14*(2), 9–38.

Kesztenbaum, L. and J.-L. Rosenthal (2017). Sewers' Diffusion and the Decline of Mortality: The Case of Paris, 1880–1914. *Journal of Urban Economics 98*, 174–186.

Kibele, E. U. (2012). *Regional mortality differences in Germany.* Springer Science & Business Media.

Krugman, P. (1993). First Nature, Second Nature, and Metropolitan Location. *Journal of Regional Science 33*(2), 129–144.

Kullback, S. and R. A. Leibler (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics 22*(1), 79–86.

Labrador, J. (2013). Une forte hétérogénéité des revenus en Ile-de-France. *Ile-de-France a la page.*

Laménie, B. (2016). Les territoires industriels face aux effets cumulés du déclin démographique et économique: quelles perspectives avec la métropolisation? L'exemple des Ardennes. *Espace populations sociétés. Space Populations Societies* (2015/3-2016/1).

Le Mée, R. (1989). Les villes de France et leur population de 1806 à 1851. *Annales de démographie historique* (1), 321–393.

Lessmann, C. (2014). Spatial Inequality and Development – Is There an Inverted-u Relationship? *Journal of Development Economics 106*, 35–51.

Lessmann, C. and A. Seidel (2017). Regional Inequality, Convergence, and its Determinants–a View from Outer Space. *European Economic Review 92*, 110–132.

Lori, L. (1972). *Les comptes régionaux des ménages de 1962 à 1967*, Volume 8 of *Collections de l'INSEE, Série R*. Institut national de la statistique et des études économiques.

Lori, L. (1973). *Les comptes régionaux des ménages de 1968 et 1969*, Volume 11 of *Collections de l'INSEE, Série R*. Institut national de la statistique et des études économiques.

Mackenbach, J. P., V. Bos, O. Andersen, M. Cardano, G. Costa, S. Harding, A. Reid, Ö. Hemström, T. Valkonen, and A. E. Kunst (2003). Widening Socioeconomic Inequalities in Mortality in Six Western European Countries. *International Journal of Epidemiology 32*(5), 830–837.

Mackenbach, J. P. and A. E. Kunst (1997). Measuring the Magnitude of Socio-economic Inequalities in Health: an Overview of Available Measures Illustrated with two Examples from Europe. *Social Science & Medicine 44*(6), 757–771.

Merton, R. K. (1968). The Matthew Effect in Science: The Reward and Communication Systems of Science Are Considered. *Science 159*(3810), 56–63.

Michaels, G., F. Rauch, and S. J. Redding (2012). Urbanization and Structural Transformation. *The Quarterly Journal of Economics 127*(2), 535–586.

Mojica, L. and J. Martí-Henneberg (2011). Railways and Population Distribution: France, Spain, and Portugal, 1870–2000. *Journal of Interdisciplinary History 42*(1), 15–28.

Oliveau, S. and Y. Doignon (2016). La diagonale se vide? Analyse spatiale exploratoire des décroissances démographiques en france métropolitaine depuis 50 ans. *Cybergeo: European Journal of Geography*.

Oswalt, P. and T. Rieniets (2006). *Atlas of Shrinking Cities*. Hatje Cantz.

Pedroncini, G., A. Corvisier, and A. Blanchard (1992). *Histoire militaire de la France. 3. De 1871 à 1940.* Presses Universitaires de France.

Piketty, T. (2001). *Les hauts revenus en France au XXème siècle.* Grasset.

Preston, S., P. Heuveline, and M. Guillot (2000). *Demography: Measuring and Modeling Population Processes.* Wiley-Blackwell.

Preston, S. H. and E. Van de Walle (1978). Urban French mortality in the XIXth century. *Population Studies 32*(2), 275–297.

Prost, A. (2008). Compter les vivants et les morts: l'évaluation des pertes françaises de 1914–1918. *Le Mouvement Social* (1), 41–60.

Puga, D. (1999). The Rise and Fall of Regional Inequalities. *European Economic Review 43*(2), 303–334.

Pumain, D. (1986). Les migrations interrégionales de 1954 à 1982: directions préférentielles et effets de barrière. *Population (French Edition) 41*(2), 378–389.

Robine, J.-M. (2001). Redéfinir les phases de la transition épidémiologique à travers l'étude de la dispersion des durées de vie: le cas de la france. *Population (french edition)*, 199–221.

Roine, J. and D. Waldenström (2008). The Evolution of Top Incomes in an Egalitarian Society: Sweden, 1903–2004. *Journal of Public Economics 92*(1-2), 366–387.

Rosental, P.-A. (2004). La migration des femmes (et des hommes) en France au XIXe siècle. *Annales de Démographie Historique 107*(1), 107–135.

Ryder, N. B. (1975). Notes on Stationary Populations. *Population Index*, 3–28.

Salehyan, I. and K. S. Gleditsch (2006). Refugees and the Spread of Civil War. *International Organization 60*(2), 335–366.

Schmeidl, S. (1997). Exploring the Causes of Forced Migration: A Pooled Time-Series Analysis, 1971–1990. *Social Science Quarterly 78*(2), 284–308.

Schulze, M.-S. (2007). Regional Income Dispersion and Market Potential in the Late Nineteenth Century Hapsburg Empire. *Economic History Working Papers 106/07.*

Schwartz, R., I. Gregory, and T. Thévenin (2011). Spatial History: Railways, Uneven Development, and Population Change in France and Great Britain, 1850–1914. *Journal of Interdisciplinary History 42*(1), 53–88.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal 27*(3), 379–423.

Sommeiller, E. and M. Price (2014). The Increasingly Unequal States of America. *Washington: Economic Policy Institute.*

Sommeiller, E., M. Price, and E. Wazeter (2016). Income Inequality in the US by State, Metropolitan Area, and County. *Economic Policy Institute, June 16.*

Talandier, M., V. Jousseaume, and B.-H. Nicot (2016). Two Centuries of Economic Territorial Dynamics: the Case of France. *Regional Studies, Regional Science 3*(1), 67–87.

Turok, I. and V. Mykhnenko (2007). The Trajectories of European Cities, 1960–2005. *Cities 24*(3), 165–182.

Vallin, J. and F. Meslé (2001). *Tables de mortalité françaises pour les XIXe et XXe siècles et projections pour le XXIe siècle*. Éditions de l'Institut National d'Etudes Démographiques.

Vallin, J. and F. Meslé (2005). Convergences and Divergences: an Analytical Framework of National and Sub-National Trends in Life Expectancy. *Genus 61*(1), 83–124.

Vallin, J., F. Meslé, G. Caselli, and V. Egidi (1988). *Les causes de décès en France de 1925 à 1978.* Ined.

Van de Walle, É. (1974). *The Female Population of France in the 19th Century: A Reconstruction of 82 Départements*. University Press.

Vlahov, D., S. Galea, and N. Freudenberg (2005). The Urban Health "Advantage". *Journal of Urban Health 82*(1), 1–4.

Waldron, H. (2007). Trends in Mortality Differentials and Life Expectancy for Male Social Security-Covered Workers, by Socioeconomic Status. *Social Security Bulletin 67*, 1.

Williamson, J. G. (1965). Regional Inequality and the Process of National Development: a Description of the Patterns. *Economic Development and Cultural Change 13*(4, Part 2), 1–84.

Wilmoth, J. R., K. Andreev, D. Jdanov, D. A. Glei, C. Boe, M. Bubenheim, D. Philipov, V. Shkolnikov, and P. Vachon (2007). Methods Protocol for the Human Mortality Database. *University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. URL: http://mortality. org [version 31/05/2007] 9*, 10–11.

Wilmoth, J. R. and S. Horiuchi (1999). Rectangularization Revisited: Variability of Age at Death Within Human Populations. *Demography 36*(4), 475–495.

Yaari, M. E. (1965). Uncertain lifetime, life insurance, and the theory of the consumer. *The Review of Economic Studies*, 137–150.

Zukin, S. (1985). The Regional Challenge to French Industrial Policy. *International Journal of Urban and Regional Research 9*(3), 352–367.

Abstract

This Ph. D. thesis has a dual purpose. First, it presents the methods used to build two new historical databases relating to départements. The first database provides the departmental lifetables for the period 1901-2014. The second database provides the departmental distributions of income over the period 1960-2014. Second, this thesis presents the first works resulting from the joint use of these two databases and other statistics. They concern both the dynamics of spatial inequalities and some specific historical events. Thus, the analysis of the spatial distribution of the population since the middle of the 19th century allows to understand the dynamics induced by the rural exodus, but also by the new trends of today's migrations. The analysis of mortality inequalities over the last 200 years shows that inequalities have fallen dramatically since the end of the 19th century, while the geography of excessmortality has changed. Finally, the analysis of spatial income inequalities reveals a continuous decline since the 1920s. This decline occurred only since 1950 if spatial inequalities are observed using asynthetic indicator of welfare, combining both mortality inequalities and income inequalities. The thesis ends with the analysis of internal migrations during the Second World War : these migrations were massive, and clearly oriented towards the free zone. These results testify both to the impact of this event on French demography, and to the quest for freedom of the French of that time, little hampered by the demarcation line.

Résumé

Cette thèse a un double objectif. En premier lieu, elle présente les méthodes ayant permis de construire deux bases de données historiques relatives aux départements français. La première met à disposition les tables de mortalité départementales sur la période 1901-2014. La seconde permet de disposer des distributions départementales de revenu sur la période 1960-2014. En second lieu, cette thèse présente les travaux issus de l'utilisation conjointe de ces deux bases de données et d'autres statistiques : ils concernent aussi bien les dynamiques longues des inégalités spatiales que certains évènements historiques. Ainsi, l'analyse de la répartition spatiale de la population depuis le milieu du 19ème siècle permet de comprendre à la fois la dynamique induite par l'exode rural, mais aussi par les nouvelles tendances des migrations d'aujourd'hui, différentes selon les âges. L'analyse des inégalités de mortalité depuis 200 ans montre quant à elle que les inégalités ont largement baissé depuis la fin du 19ème siècle, alors que la géographie de la surmortalité a profondément changé. Enfin, l'analyse des inégalités spatiales de revenus révèle une baisse continue des inégalités depuis les années 1920, baisse qui n'intervient que depuis 1950 si l'on introduit les inégalités de mortalité dans un indicateur synthétique de bien-être. La thèse se conclut par l'analyse des migrations internes durant la seconde guerre mondiale : leur caractère à la fois massif et à destination de la zone libre témoigne aussi bien de l'impact qu'a eu cet évènement sur la démographie française que de la formidable quête de la liberté des français de l'époque, peu entravée par la ligne de démarcation.