

# Micro-simulation sur le cycle de vie : le modèle TaxIPP-Life

Béatrice Boutchenik et Alexis Eidelman  
IPP

5 mai 2013

Document extrêmement préliminaire. Ne pas diffuser, ne pas citer.

Ce document présente le modèle TaxIPP-Life. La construction de ce modèle a débuté il y a moins d'un an au moment de la rédaction de ce rapport. Le modèle est encore dans une phase de développement. Les résultats du modèle correspondent pour l'instant plus à des exercices de test qu'à des statistiques.

# Table des matières

<b>Introduction</b>	<b>3</b>
Motivations . . . . .	3
Les difficultés d'une étude sur le cycle de vie . . . . .	8
<b>1 Construction de la base de données</b>	<b>9</b>
1.1 Appariement statistique des données EIC-DADS avec l'Enquête Patrimoine . . . .	9
1.1.1 Mise en forme des fichiers EIC-DADS-UNEDIC . . . . .	9
1.1.2 Mise en forme de l'Enquête Patrimoine . . . . .	10
1.1.3 Appariement des deux bases . . . . .	11
1.2 Trajectoires salariales . . . . .	12
1.2.1 Imputation des années manquantes . . . . .	12
1.3 Résultats préliminaires : description des trajectoires salariales . . . . .	13
1.3.1 Volatilité des revenus sur le cycle de vie . . . . .	13
<b>2 Simulation sur le cycle de vie</b>	<b>16</b>
2.1 Quelques mots de technique . . . . .	16
2.1.1 La base de données initiales . . . . .	16
2.1.2 La fermeture de l'échantillon . . . . .	17
2.1.3 Étendre l'échantillon . . . . .	18
2.1.4 Choix de langage de programmation . . . . .	19
2.2 Les étapes de la simulation dynamique . . . . .	19
2.2.1 Démographie . . . . .	19
2.2.2 Logement . . . . .	20
2.2.3 Marché du travail . . . . .	21
2.2.4 Épargne et consommation . . . . .	22
2.2.5 Calcul de la législation . . . . .	22
2.3 Projets futurs . . . . .	22
2.3.1 Simulation par mois . . . . .	22
2.3.2 Transferts intergénérationnels . . . . .	23
2.3.3 Comportements . . . . .	23
2.3.4 Autres enquêtes . . . . .	24
<b>3 Résultats préliminaires : simulation des allocations familiales (...et de ??)</b>	<b>24</b>

## Introduction

Dans cette introduction nous détaillerons les motivations du projet TaxIPP-Life puis après avoir souligné les principales difficultés associées aux analyses sur le cycle de vie. Nous annoncerons le plan du présent document et quelles sont les solutions pour répondre à ces difficultés.

## Motivations

L'objet du projet TaxIPP-Life, débuté en septembre 2012, est la réalisation d'études sur la population française portant sur l'ensemble du cycle de vie. Le regard que l'on porte sur les inégalités peut se voir modifié dans cette optique. En effet, Lorsqu'on s'intéresse à des questions empiriques touchant aux inégalités et à la redistribution, dont les réponses dépendent en grande partie de la distribution effective des revenus, le fait de se concentrer sur une année donnée seulement a de fortes chances de mener à une vision erronée des phénomènes étudiés. Premièrement, les revenus des individus peuvent en effet être amenés à varier de façon considérable au cours de leur vie, et la mobilité de ceux-ci dans la distribution générale des revenus aura des conséquences en termes de mesure des inégalités, et donc dans la réduction de celles-ci par des transferts dont le ciblage ne dépend en général que de l'année en cours. Deuxièmement, l'impact redistributif des différents éléments du système fiscal-social est moins bien appréhendé lorsque l'on fait une étude en coupe. En effet, lorsque l'on étudie une année donnée, on manque le fait qu'une partie des prélèvements sur les revenus d'un individu lui reviendront ou lui ont été avancé à d'autres moment de sa vie. Pareillement, une partie des prestations sont financée par l'individu au cours de sa vie. Une étude sur le cycle de vie pourra séparer une partie intra-personnelle de la redistribution (au cours du temps) et une partie inter-personnelle qui correspond à une sorte de vraie redistribution des revenus de certains vers d'autres. L'analyse sur cycle de vie peut aussi tenir compte des transferts dit assurantiels contrairement aux études en coupe qui sont contraintes à les laisser hors champs.

Les revenus perçus par les individus sont susceptibles d'évoluer au cours du temps selon deux mécanismes principaux. D'une part selon une perspective de cycle de vie traditionnelle dans laquelle, après une période de formation, les individus ont des revenus faibles en début de vie, puis des revenus de plus en plus importants durant leur carrière et enfin une retraite pendant laquelle leurs revenus sont diminués. D'autre part, des chocs, parfois imprévisibles, chômage, maladie, d'inactivité, etc., peuvent affecter temporairement les revenus. Ces deux facteurs font que les individus restent rarement toute leur vie au même point de la distribution globale des revenus. [Slemrod, 1992] montre par exemple dans le cas des États-Unis qu'entre 1980 et 1986, lorsque l'on considère une année donnée, plus de 20 % des individus situés dans le décile de revenus le plus élevé n'y étaient pas l'année précédente. Ce chiffre est de 33 % lorsqu'on considère le percentile le plus haut, entre 1982 et 1985. Il met par ailleurs en évidence le fait que pour les catégories de revenus les plus faibles en 1983 (allant jusqu'à 25000 \$ annuels), le revenu moyen sur la période 1979-1985 est systématiquement plus élevé que le revenu pour la

seule année 1983. Ceci est d'autant plus vrai que la classe de revenu est faible : les individus dont les revenus en 1983 étaient compris entre 0 et 5000 \$ ont un revenu moyen de 3063 \$ en 1983, contre un revenu moyen de 7395 \$ sur la période considérée.

[Bjorklund, 1993] examine dans le cas de la Suède la relation entre les revenus en coupe et sur le cycle de vie, de 1951 à 1989. Les corrélations entre ces deux revenus sont globalement faibles. Elles sont même presque nulles, lorsqu'on considère les individus à un âge situé entre 25 et 30 ans, alors qu'elles sont relativement fortes pour des groupes plus âgés. Le revenu annuel serait donc un indicateur particulièrement mauvais du revenu sur le cycle de vie pour les jeunes.

Cette variabilité des revenus au cours de la vie des individus porte un certain nombre de conséquences quant à la mesure des inégalités. Si les individus se situent à différents points de la distribution de revenus à différents moments de leur vie, il en découle que la vision des inégalités prise en coupe sera exagérée par rapport aux inégalités de revenu sur le cycle de vie. Considérons par exemple une économie d'individus parfaitement identiques et ayant le même profil de revenus au cours de leur cycle de vie (et ce dans un monde sans croissance réelle des revenus au cours du temps), mais qui diffèrent par leur âge : on constate l'existence d'inégalités pour une année donnée, alors même que chaque individu aura reçu à chaque âge un revenu réel identique à celui reçu par les autres au même âge. [Fitzgerald and Maloney, 1990] ont ainsi calculé que le coefficient de Gini calculé pour les États-Unis en l'année 1979 se voyait réduit de 19,1 % après avoir contrôlé par le mouvement d'ascension des revenus selon l'âge, à travers les cohortes.

D'autres études empiriques ont étudié les inégalités de revenus considérés sur des périodes plus ou moins longues, en prenant donc en compte à la fois l'évolution « naturelle » des revenus sur le cycle de vie et les chocs transitoires pouvant affecter ceux-ci. [Slemrod, 1992] rapporte les résultats de [Shorrocks, 1978] et de [Bemis and Morgan, 1975] qui montrent que l'extension à deux années ou plus de la période sur laquelle les inégalités sont étudiées change très peu la mesure de celles-ci. Il s'intéresse lui-même à la fraction de revenus détenue par les 1 %, 5 % et 10 % mieux lotis de la population, selon qu'on considère une année donnée ou la période allant de 1979 à 1985. Celle-ci passe respectivement de 8.39 à 7.64 %, de 19.94 à 18.70 % et de 29.96 à 28.50 % pour les trois catégories d'individus considérées.

Au contraire, selon [Pettersson and Pettersson, 2007] qui utilisent des données simulées, le coefficient de Gini diminue de 60 % si l'on considère tout le cycle de vie, par rapport à la même mesure pour une année donnée. citeBjorklund1993 estime que cette diminution est de l'ordre de 35 à 40 %, en considérant des données suédoises effectivement observées entre 1951 et 1989, soit une période bien plus longue que celle étudiée par Slemrod<sup>1</sup>. Il est logique que la durée de la période considérée ait un effet important sur les conclusions auxquelles on parvient en termes d'inégalités et de redistribution, ce sur quoi on reviendra par la suite.

Les études portant sur les inégalités annuelles de revenus mélangent les variations de revenus

---

1. Il est utile de préciser ici qu'une révision à la baisse de la mesure des inégalités n'implique pas de porter une attention réduite aux bas revenus, qui même s'ils sont temporaires correspondent quand même à des situations défavorables dans l'immédiat, en l'absence d'accès parfait aux marchés financiers, et doivent être corrigés

au cours du temps pour un même individu et les inégalités entre revenus individuels totaux calculés sur la période considérée, en leur accordant le même poids. Même si la situation à instant donné n'est pas sans intérêt, en particulier si l'on pense aux personnes ayant des revenus faibles temporairement dans un contexte où l'accès au marché financier n'est pas parfait, ne pas tenir compte de l'âge où des revenus sur l'ensemble du cycle de vie est une limite pour ces études. Les analyses portant au contraire sur une période plus longue permettent au contraire de dissocier, au sein d'une mesure totale des inégalités, une partie tenant réellement aux inégalités entre individus (inégalités «inter-individuelles») et une partie tenant à la variabilité des revenus sur le cycle de vie (inégalités «intra-individuelles»).

Toutes les études montrent, avec des ampleurs différentes, que la mesure des inégalités mesurée en coupe est surévaluée par rapport à son estimation sur le cycle de vie. Si l'on croit que la consommation peut être lissée au cours de la vie et que le revenu, c'est bien un revenu moyen sur plusieurs années dont il faut tenir compte pour estimer le niveau de vie d'un individu.

La question du patrimoine est elle aussi primordiale dans l'étude des inégalités. Il est en effet raisonnable de croire qu'un individu ayant un capital (éventuellement un capital attendu à travers un héritage) n'aura pas la même attitude de consommation qu'une personne qui n'en a pas, ce que l'on traduit par des niveaux de vie différents. Si on pense à deux étudiants, l'un aidé par ses parents, l'autre non, celui aidé par ses parents est généralement considéré comme étant plus aisé que l'autre. Pourtant, si pour subvenir à ses besoins le second travaille en même temps que ses études, il sera paradoxalement considéré comme étant plus riche. De façon générale, la «richesse» d'un individu est un mélange de ses revenus et de son patrimoine, les études en coupe sont encore limitées parce que le patrimoine, parce qu'il est parfois peu liquide ou précisément parce qu'il ne peut s'expliquer que par des considérations sur cycle de vie, est exclu de l'analyse. Le revenu potentiel qu'il représente n'est pourtant pas négligeable et la différence entre un propriétaire et un locataire se situe à la fois dans le loyer (et ce point devrait toujours être réglé dans les études d'inégalités) et à la fois dans le revenu actualisé que pourrait représenter la vente du logement.

Le premier point pour justifier une approche sur cycle de vie est donc que l'on peut mieux appréhender la situation des individus. Un deuxième est que si l'on veut étudier la progressivité des transferts ou du système socio-fiscal dans son ensemble, une étude en coupe souffre de limitation importante.

Le constat porté sur le revenu en coupe de individu qui n'est pas suffisant pour appréhender leur richesse tient aussi pour la redistribution. Une analyse en coupe, sans être dénuée d'intérêt, ne permet pas de réaliser qu'une partie des transferts correspondent moins à de la redistribution qu'à un lissage forcé des revenus, ou à une consommation forcée (d'éducation par exemple mais aussi d'assurance) dont l'individu bénéficie personnellement. Si on reprend l'exemple théorique d'une population d'individus parfaitement identiques et ayant le même profil de revenus au cours

de leur cycle de vie et qu'un impôt est prélevé pour financer un certain transfert, par exemple un impôt sur le revenu pour financer des retraite ou de l'éducation, une année donnée, il y aura un transfert des actifs vers les jeunes et les retraités mais dans cette population, toutes les carrières sont les mêmes si bien que l'ensemble des prélèvements sur les revenus d'un individu lui reviendront sous forme de retraite ou de dépense d'éducation.

Empiriquement, [Bengtsson et al., 2012] montrent ainsi que la progressivité du système fiscal suédois sur le cycle de vie est plus faible que sa progressivité pour une année donnée, ce pour presque toutes les années de la période 1968-2009. Ainsi le système serait presque proportionnel et aurait un effet redistributif de seulement quelques points de pourcentage, réduisant le coefficient de Gini de 10 %. [Bjorklund, 1993] met quant à lui en évidence une réduction de celui-ci de 20 % sur la période 1978-1990, pour la Suède également, ce qui montre l'importance de l'étendue de la période étudiée. Pour les Etats-Unis, [Slemrod, 1992] trouve un effet redistributif compris entre 4 et 6 % lorsqu'on considère le cycle de vie. A titre d'exemple, le premier décile est taxé en moyenne à hauteur de 4.6 % lorsqu'on considère une année donnée, contre 6.1 % sur le cycle de vie, le deuxième à 5.5 et 7.4 % respectivement pour une année et sur le cycle de vie. Les écarts sont bien moins grands dans le haut de la distribution. De même, citeBengtsson2012 montrent que ce sont surtout les quatre premiers quintiles qui sont affectés par la perspective de cycle de vie, alors que celle-ci ne change pas de beaucoup le taux de taxation effectif pour le quintile de revenus le plus élevé.

Ces travaux amènent donc bien à nuancer la redistribution effectivement opérée entre individus par les systèmes fiscal-socials existants. Celle-ci ne serait donc pas uniquement « inter-personnelle », mais également « intra-personnelle ». Ainsi [Pettersson and Pettersson, 2007] estiment qu'entre 18 et 32 % de la redistribution effectuée par le système de redistribution suédois (taxes et prestations monétaires et non-monétaires) est effectivement de la redistribution entre individus, le reste correspondant à de la redistribution intra-personnelle qui est donc autofinancée : environ 7 à 8 couronnes suédoises sur 10 reçues sont payées, en moyenne, par les individus eux-mêmes à un autre moment. Dans le cas de la Suède, l'effet principal du système de redistribution serait donc un transfert de ressources au sein du cycle de vie. Les auteurs rapportent les résultats similaires de [Hussenius and Selen, 1994] qui pour ce même pays estiment que le taux de redistribution inter-personnelle se situe entre 24 et 32 %. Étudiant le cas de l'Australie, [Falkingham et al., 1996] situent ce chiffre entre 48 et 63 %, et entre 29 et 38 % pour le Royaume-Uni. Il est par ailleurs estimé à 45 % pour l'Irlande et à 24 % pour l'Italie [?].

La part de redistribution intra-personnelle des systèmes fiscal-socials des différents pays pour lesquels elle a été mesurée apparaît donc substantielle. Celle-ci reflète une logique d'assurance contre les périodes du cycle de vie où les individus voient leurs revenus baisser ou certains types de dépenses augmenter, et ce de façon plus ou moins prévisible. [Varian, 1980] a ainsi souligné dès 1980 que la taxation peut être considérée comme une assurance sociale, ayant pour

rôle d'égaliser le revenu et la consommation sur le cycle de vie. L'ensemble du système redistributif participe donc potentiellement de la fonction assurantielle de l'Etat, dans un monde où les marchés financiers ne sont pas toujours complets. On aura pu noter dans le paragraphe précédent que les taux de redistribution inter-personnelle (rapportée à la redistribution totale) sont relativement plus élevés dans les pays anglo-saxons qu'en Suède, par exemple. Ceci est à rapprocher du constat de [Lindert, 2004] selon lequel les pays ayant un ratio élevé de dépenses publiques rapportées au PIB ont également un système fiscal plus proportionnel, comparative-ment aux pays anglo-saxons. Dans ces derniers, la fonction d'assurance est plutôt assurée par les marchés, ce qui laisse à l'Etat celle d' « assistance ». Ceci est également cohérent avec la taxonomie canonique de l'Etat-Providence ([Esping-Andersen, 1990]), dans laquelle les Etats scandinaves sont caractérisés par un haut niveau d'universalité et une dépendance limitée aux marchés et à la famille.

La distinction entre redistribution inter-personnelle et redistribution intra-personnelle (ou assurantielle) revêt donc une grande importance dans le cadre du débat public. D'une part lorsqu'il s'agit de comparer par exemple les taux de prélèvements obligatoires entre différents pays : il est nécessaire de garder à l'esprit que les écarts observés peuvent s'expliquer pour une large partie par le caractère public ou privé des fonctions d'assurance. D'autre part, cette distinction est cruciale dans la réflexion menée sur l'équité des systèmes redistributifs : non seulement la redistribution effective entre individus doit être corrigée de la surestimation qui intervient lorsqu'on se restreint à une seule année, mais il est également important de prendre en compte les conséquences en termes de bien-être de la redistribution intra-personnelle, conséquences qui peuvent être inégalement réparties au sein de la population. C'est ce que montrent [Hoynes and Luttmer, 2011] dans le cas des Etats-Unis : la valeur de la composante assurantielle de la redistribution est positive pour l'ensemble de la population, mais elle est d'autant plus grande que le revenu est élevé.

Tous les éléments des systèmes de redistribution ne contribuent pas de façon égale aux composantes de redistribution intra- et inter-personnelle de ceux-ci. [Björklund and Palme, 1997] montrent ainsi que l'impôt sur le revenu suédois agit principalement sur les inégalités entre individus (considérés sur le cycle de vie), alors que l'allocation familiale universelle a un effet à la fois sur celles-ci et sur les inégalités au sein du cycle de vie pour chaque individu. Ceci est tout à fait logique puisque les auteurs prennent en compte la consommation des enfants à travers une échelle d'équivalence, et que le but des allocations familiales est d'apporter un soutien économique dans les périodes de plus lourdes responsabilités familiales. L'allocation logement suédoise, accordée sous condition de ressources contrairement aux allocations familiales, possède elle aussi un effet à la fois sur les inégalités inter- et intra-personnelles. Les auteurs montrent par ailleurs que la volatilité des revenus est la plus forte pour les individus ayant les plus bas revenus totaux sur le cycle de vie, et que c'est donc sur leurs revenus que le système redistributif exerce le plus fort lissage. De même, la redistribution intra-personnelle que mettent en évidence [Bengtsson et al., 2012] reflète d'après eux la nature transitoire de la perception de bas revenus

plutôt que la nature transitoire des hauts revenus. La perspective de cycle de vie serait donc particulièrement importante lorsqu'on considère les éléments redistributifs touchant les individus à bas revenus : c'est pour ces composantes qu'on pourrait observer l'effet assurantiel le plus fort relativement à l'effet réellement redistributif.

## Les difficultés d'une étude sur le cycle de vie

L'étude sur le cycle de vie exige, et c'est une lapalissade, des données sur l'ensemble de ce cycle de vie. La base de données souhaitable serait un panel sur l'ensemble du cycle de vie contenant toutes les informations utiles sur le patrimoine, les revenus, les statuts conjugaux, la consommation, etc. Si des données proches de cet idéal existe dans les pays scandinaves, ce n'est pas le cas en France. En interrogeant un individu une année donnée, pour connaître son cycle de vie s'expose à deux problèmes. Le plus évident est que lorsque l'on interroge un individu, il ne connaît pas son avenir. Le modèle TaxIPP-Life contient un module réalisant une projection à un niveau individuel du futur. Ceci sera développé dans la deuxième partie de ce document. La première s'attardera sur un point moins évident qui est que le passé des individus n'est pas non plus connu parfaitement. En dehors des effets de mémoire qui font qu'une partie du passé est oubliée, aucune enquête ne demande à un individu de détailler l'ensemble de ses revenus passés. Des données administratives semblent être la seule solution pour connaître par exemple la trajectoire professionnelle des individus. Mais les données administratives ne sont pas non plus une panacée, elles ne donnent en général pas d'information sur la vie matrimoniale des personnes, or, savoir qui vit avec qui est déterminant pour étudier la consommation et les niveaux de vie. Enfin, une troisième voie serait l'utilisation d'un panel qui suivrait les individus depuis leur naissance ou du moins pendant un certain temps, toutefois une telle base de données, avec des informations assez riches pour pouvoir appliquer la législation socio-fiscal n'existe pas. Nous avons donc procédé à un matching statistique de données d'enquête et de données administratives. La première partie de document décrira ce matching avec entre les données de l'enquête patrimoine et les données de l'échantillon EIR-EIC.

Enfin, la troisième partie de ce document présentera des résultats en ce centrant sur les prestations. Si l'objectif du projet TaxIPP-Life est, à terme, de simuler l'ensemble du système de redistribution français sur le cycle de vie. Il paraît toutefois intéressant de se concentrer dans un premier temps sur les prestations sociales. Celles-ci opèrent en effet une grande partie de la redistribution observée en coupe en France. D'après [Duval et al., 2011], elles contribuent pour deux tiers à la réduction des inégalités en France en 2011, contre un tiers pour les prélèvements. De plus, elles sont l'instrument le plus susceptible d'avoir un impact sur les individus à bas revenus<sup>2</sup>, dont on a vu qu'ils connaissent la volatilité de revenus sur le cycle de vie la plus élevée. C'est donc la redistributivité des prestations sociales qui risque d'être la plus affectée par la perspective de cycle de vie. Enfin, dans le cadre du débat actuel portant sur l'équilibre de la

---

2. Le rapport [Bourguignon, 1998] montre ainsi que les instruments dominants de la redistribution sont les transferts sous condition de ressources dans le bas de la distribution, et la progressivité de l'impôt sur le revenu dans le haut de la distribution.



branche famille de la Sécurité sociale et après la publication du rapport [Fragonard et al., 2013], un éclairage dans une perspective de cycle de vie sur certaines propositions de réformes, notamment en ce qui concerne les allocations familiales, paraît particulièrement intéressant.

## 1 Construction de la base de données

### 1.1 Appariement statistique des données EIC-DADS avec l'Enquête Patrimoine

Un des intérêts majeurs de TaxIPP-Life réside dans l'utilisation de trajectoires salariales passées effectivement observées, plutôt que simulées. Les données administratives des EIC et DADS ont été utilisées à cette fin. Celles-ci ont été couplées à l'enquête Patrimoine afin de disposer d'un certain nombre d'informations, notamment au niveau du ménage plutôt que de l'individu. Un appariement statistique a donc été effectué entre les deux sources de données : appariement statistique exact sur un certain nombre de variables considérées au moment de l'enquête (sexe, âge, PCS à 1 chiffre, tranche de revenus salariaux et de remplacement), puis appariement avec le plus proche voisin, en particulier en termes de trajectoire vis-à-vis de l'emploi.

Il a été dans un premier temps nécessaire de manipuler les fichiers EIC-DADS (et UNEDIC) d'une part et l'enquête Patrimoine d'autre part afin de les rendre plus exactement comparables quant aux variables utilisées pour l'appariement statistique : il faut en effet que ces variables soient définies de la façon la plus proche possible dans une base et dans l'autre. En particulier, il a fallu utiliser l'information des fichiers EIC-DADS-UNEDIC afin de retracer les trajectoires vis-à-vis de l'emploi, selon les mêmes modalités que celles renseignées dans le calendrier rétrospectif de l'Enquête Patrimoine, calendrier renseignant les changements entre des situations ayant duré au moins un an.

#### 1.1.1 Mise en forme des fichiers EIC-DADS-UNEDIC

On conserve uniquement les individus nés à partir de 1942, les données concernant les générations 1934 et 1938 n'étant pas complètes.

**Reconstitution des trajectoires vis-à-vis de l'emploi** Les observations du fichier DADS correspondent au croisement individu x année x entreprise. A partir de ce fichier et pour les années 1976 à 2001, on détermine le statut de l'individu au 1er janvier de chaque année, en termes d'emploi à temps plein ou temps partiel. Les données du fichier EIC caisse x individu x année permettent par ailleurs de croiser cette information avec le fait que le salarié était dans le public, le privé ou encore était indépendant pour l'année donnée. Ces sources permettent de retracer les modalités suivantes de la variable de statut vis-à-vis de l'emploi de l'Enquête Patrimoine : salarié du public à temps complet, salarié du public à temps partiel, salarié du privé à temps complet, salarié du privé à temps partiel et à son compte . Lorsqu'il y a cumul

de plusieurs emplois au premier janvier de l'année, que ceux-ci soient à temps plein ou temps partiel, on considère que l'individu travaille à temps plein. Enfin, l'appartenance au public ou à la catégorie des indépendants étant obtenue dans l'EIC grâce à l'appartenance à différentes caisses - appartenance dont seule l'année est connue et non les dates exactes au cours de cette année -, on a 2116 individus qui pour au moins une année sont affiliés à la fois à une caisse du public et à une caisse d'indépendants. On attribue aléatoirement un des deux statuts à ces individus pour l'année donnée.

Le fichier UNEDIC permet quant à lui de retracer pour les années 1984 à 2001 et grâce au type d'allocation perçue le fait qu'au 1er janvier de chaque année l'individu soit en situation de chômage indemnisé ou non (chômage indemnisé seulement jusqu'en 1992), de formation et de préretraite. S'il y a un cumul de plusieurs allocations dont une de préretraite ou de formation au 1er janvier de l'année considérée, on classe l'individu comme étant plutôt en préretraite ou formation. Lorsqu'on a à la fois un statut provenant du fichier DADS et un statut provenant du fichier UNEDIC pour le même individu et la même année, on conserve le statut DADS. L'information provenant du fichier EIC caisse x individu x année, complétée par les dates de sortie des fichiers EIC-DADS-UNEDIC, nous permet de retracer les départs à la retraite. Enfin, le calendrier rétrospectif de l'enquête Patrimoine comprend une modalité "succession de courtes périodes (inférieures à un an) d'emploi et de chômage", qu'il serait dommage d'assimiler aléatoirement à de l'emploi ou du chômage, celle-ci nous renseignant sur un type de situation particulière vis-à-vis de l'emploi. De telles périodes ont ainsi été repérées dans le fichier DADS.

**Revenus salariaux et de remplacement pour l'année 2001** On additionne les salaires nets renseignés dans les DADS et les salaires nets des fonctionnaires de l'Etat donnés dans la base EIC caisse x individu x année. On prend enfin en compte les revenus de remplacement, que l'on divise par 0,7 afin de comparer des salaires de référence et non des revenus effectifs, qui nous intéresseraient moins ici : il serait par exemple peu souhaitable d'apparier un individu gagnant 2000 euros mensuels avec un autre gagnant généralement 3000 euros environ, mais qui l'année considérée se trouve être en situation de chômage et donc le revenu de remplacement est donc plus proche de 2000 euros.

### 1.1.2 Mise en forme de l'Enquête Patrimoine

L'appariement se fera naturellement lorsque l'on disposera du fichier EIC-DADS 2010, mais on matche pour l'instant l'Enquête Patrimoine 2009-2010 et l'EIC 2001 en alignant la première sur l'EIC, et en considérant donc la situation d'un individu pour l'année 2010 dans l'Enquête Patrimoine comme correspondant à celle pour l'année 2001 de l'EIC. Ainsi, on utilise pour l'appariement selon les trajectoires des dates retardées de 9 années pour l'Enquête Patrimoine, ce qui revient à comparer des situations vis-à-vis de l'emploi pour un âge donné. On déflate par ailleurs les revenus salariaux et de remplacement de l'Enquête Patrimoine 2010 par la croissance nominale observée des salaires afin qu'ils soient comparables à ceux provenant des fichiers EIC-

DADS-UNEDIC en 2001. De même que pour les données issues du fichier EIC, on multiplie les revenus de remplacement par un facteur de 0.7 avant de les additionner aux revenus salariaux.

Les données EIC étant disponibles pour une génération sur quatre, on regroupe également les individus de l'Enquête Patrimoine en catégories correspondant à quatre années de naissance : les individus de l'Enquête Patrimoine nés entre 1941 et 1944 seront par exemple appariés avec la génération née en 1942 du fichier EIC, ceux nés entre 1945 et 1948 avec la génération 1946, et ainsi de suite. L'échantillon est donc composé d'individus nés entre 1941 et 1972.

On conserve pour l'appariement statistique uniquement les individus qui au moment de l'enquête ne sont pas agriculteurs ou indépendants, et qui dans leur trajectoire rétrospective de l'Enquête Patrimoine ont déclaré au moins une période comme salarié (à temps plein ou temps partiel, dans le public ou le privé), comme au chômage ou en alternance de courtes périodes d'emploi et de chômage, entre 1984 et 2001. Pour ceux qui n'ont pas connu de tels épisodes sur cette période, on matche séparément ceux qui en ont connu au moins une sur la période 1976 à 1983. Les autres ne sont donc pas appariés et se verront attribuer des revenus salariaux nuls entre 1976 et 2001, ce qui est cohérent leur déclaration dans l'enquête Patrimoine. On dissocie en effet 1976-1983 et 1984-2001 car la variable de statut vis-à-vis de l'emploi ne peut être définie de la même manière sur ces deux périodes, les données UNEDIC n'étant pas renseignées avant 1984.

Enfin, on n'apparie pas pour l'instant les individus à la retraite<sup>3</sup>.

### 1.1.3 Appariement des deux bases

L'appariement est effectué de façon exacte sur les variables de sexe, d'âge par tranches de quatre ans, de PCS à un chiffre et de revenus salariaux et de remplacement par tranches au nombre de 12, entre les données de l'année 2001 dans le fichier EIC-DADS et celles de l'année 2010 pour l'Enquête Patrimoine (ces dernières étant ré-évaluées pour les revenus salariaux et de remplacement). Au sein d'une cellule d'individus possédant les mêmes caractéristiques, on recherche alors un plus proche voisin quant à la trajectoire en termes de statut entre 1984 et 2001, sauf pour les individus n'ayant pas eu de période salariée ou de chômage entre 1984 et 2001, et aux revenus salariaux et de remplacement.

La distance entre trajectoires est calculée en spécifiant des coûts de suppression d'un statut - il est ainsi coûteux de passer d'une trajectoire à la trajectoire identique, avec simplement une période en moins ou en plus - et de substitution d'un statut à un autre. Le coût de suppression est le même quel que soit le statut pendant la période supprimée, mais les coûts de substitution varient selon le statut que l'on substitue à un autre statut donné : on doit ainsi définir une matrice de coûts (cf. infra). On n'autorise pas les statuts manquants au cours ou à la fin d'une trajectoire : une période manquante est catégorisée comme "Inactif et autres". On autorise par contre les statuts manquants au début des trajectoires, c'est-à-dire avant l'entrée dans une des

---

3. On ne dispose pas de suffisamment d'individus à la retraite dans l'échantillon EIC 2001 (puisque l'on ne prend en compte que les générations nées à partir de 1942).

bases de données, si celle-ci a lieu après 1976. Une trajectoire considérée est ainsi plus courte si elle comporte des statuts manquants au début, mais cela n'est pas vrai si elle comporte des statuts manquants au milieu ou à la fin.

**Définition de la matrice de coûts pour la distance entre trajectoires** La distance entre deux statuts est définie comme étant la différence en valeur absolue entre les revenus (d'activité et de remplacement) médians parmi les sous-groupes se trouvant dans chacun des deux statuts, en 2010. On choisit de considérer des distances en termes de salaire puisque c'est finalement l'adéquation de la trajectoire salariale de l'EIC avec la trajectoire salariale réelle connue par l'individu de l'Enquête Patrimoine qui nous importe. Il est évident que les distances entre statuts en termes de revenus salariaux et de remplacement ont pu varier entre 1984 et 2010, mais il nous faut définir une unique matrice de coûts de substitution.

**Poids relatifs de la distance entre les salaires en 2001 et de la distance entre trajectoires** Différents essais ont été effectués pour l'appariement afin de déterminer l'ordre de grandeur du poids qu'il faut attribuer à la distance entre les trajectoires relativement à celle entre les revenus salariaux et de remplacement en 2001. Lorsque l'on augmente le poids relatif attribué à la distance entre trajectoires, même considérablement, le pourcentage de statuts qui ne correspondent pas pour une période donnée entre ce qui est déclaré dans l'Enquête Patrimoine et ce qui est connu à travers l'EIC ne diminue pas beaucoup. Au contraire, la distance moyenne entre les revenus dans l'Enquête Patrimoine et dans l'EIC augmente fortement. On choisit donc d'accorder un poids relativement élevé à la distance entre les revenus en 2001, puisque cela ne détériore pas grandement la qualité de l'appariement en termes de trajectoires de statuts, tout en améliorant assez nettement celle en termes de revenus salariaux et de remplacement en 2001.

## 1.2 Trajectoires salariales

### 1.2.1 Imputation des années manquantes

**Imputation des années manquantes pour la Fonction Publique d'Etat** Les années 1979, 1981 et 1987 sont manquantes en totalité pour la Fonction Publique d'Etat. Dans le fichier EIC, les salaires de ces années ont été imputées en étudiant la probabilité d'être entré dans la fonction publique l'année  $N$  plutôt que l'année  $N + 1$  (si l'individu n'était pas déjà dans la base l'année  $N - 1$ ). Cette imputation de la présence dans la fonction publique est conservée, mais plutôt que de reprendre simplement le salaire de l'année précédente, on fait en sorte qu'il y ait une augmentation progressive entre les salaires des années  $N - 1$ ,  $N$  et  $N + 1$ . Plus précisément, on veut qu'en moyenne :

$$\begin{aligned} W_{N+1} &= W_N \cdot g_{N+1} \cdot t \\ W_N &= W_{N-1} \cdot g_N \cdot t \end{aligned}$$

avec  $g$  la croissance nominale des salaires dans le privé et semi-public (données INSEE) et  $t$  un facteur de croissance spécifique à la fonction publique et constant entre  $N - 1$  et  $N + 1$ , et que l'on calcule donc comme  $t = \sqrt{\frac{W_{N+1}}{g_{N+1} \cdot W_{N-1} \cdot g_N}}$ .

On choisit ensuite d'appliquer ce facteur (multiplié par la croissance nominale) au salaire de l'année  $N - 1$  lorsque celui-ci existe, et au salaire de l'année  $N + 1$  dans le cas contraire, et ce si l'individu est présent dans la fonction publique l'année  $N$ . Les salaires des années 1976 et 1977 sont connus dans les DADS mais pas dans les EIC pour la Fonction Publique d'Etat (l'appartenance au fichier de paie des agents de l'Etat est par contre connue). On impute pour l'année 1977 le salaire de 1978 déflaté par la croissance réelle des salaires dans le privé et le semi-public si l'individu appartient à la Fonction Publique cette année-là, et de même pour 1976 en utilisant l'année 1977.

**Imputation des années manquantes pour le secteur privé** Les années 1981, 1983 et 1990 sont entièrement manquantes dans les DADS. On estime les revenus salariaux pour chacune de ces années grâce à des modèles tobit. Pour l'année 1981, on estime tout d'abord un modèle tobit en régressant les salaires nets totaux (issus de l'emploi privé) pour l'année 1980 sur les salaires (issus du privé) de 1979 et 1978, l'âge et l'âge au carré en 1980, le sexe, et la présence dans une caisse du secteur privé en 1980 (connue grâce au fichier EIC), ce grâce à la procédure de Heckman. La sélection se fait sur les mêmes variables, avec en supplément le statut vis-à-vis de l'emploi en 1978 et 1979. Pour cette estimation, on utilise uniquement les individus pour lesquels il existe au moins un salaire strictement positif pour les années 1978, 1979 et 1982. Le salaire comme variable explicative est renseigné de façon catégorielle, par l'appartenance à l'un des neuf premiers déciles ou aux catégories P90-95, P95-99, P99-99.5, P99.5-99.9, P99.9-99.95, P99.95-99.99, et P99.99-100. Le fait de prendre en compte de façon fine l'appartenance au haut de la distribution les années passées permet par la suite de simuler des salaires suffisamment élevés pour l'année en cours. La variable de statut vis-à-vis de l'emploi est celle qui a été construite pour l'appariement, et qui reprend une partie des catégories de la variable CYACT de l'enquête Patrimoine. Une fois cette estimation effectuée, on effectue la prédiction des salaires nets de l'année 1981 ainsi que de la probabilité d'avoir un salaire positif, à partir des salaires des années 1980 et 1979, de l'âge et de l'âge au carré en 1981, etc. On n'applique cela qu'aux individus pour lesquels au moins l'un des salaires de 1979, 1980 ou 1982 est strictement positif. On impute un salaire nul aux individus dont les salaires de ces années sont tous nuls.

Pour l'année 1990, on transpose la même procédure en estimant tout d'abord un modèle tobit régressant les salaires de 1989 sur les données de 1987, 1988 et 1991 (et 1989 pour les variables d'âge et d'appartenance à une caisse du secteur privé), puis en effectuant la prédiction des salaires de 1990 à partir des données concernant les années 1988, 1989 et 1992. On applique la même procédure pour l'année 1983. Il n'est pas possible d'utiliser l'année 1981 pour l'estimation et/ou la prédiction, cette année ayant un statut particulier puisqu'elle a elle-même été imputée. On estime ainsi l'effet des salaires etc. des années 1980, 1985 et 1986 sur l'année 1984, puis on obtient la prédiction des salaires de 1983 en utilisant les données de 1979, 1984 et 1985.

## 1.3 Résultats préliminaires : description des trajectoires salariales

### 1.3.1 Volatilité des revenus sur le cycle de vie

On considère ici les trajectoires salariales sur la période allant de 1976 à 2001, dans le but d'examiner les inégalités intra-individuelles de revenu sur une partie du cycle de vie. Pour cela, on s'inspire du travail effectué par [Björklund and Palme, 1997] dans le cas de la Suède. Les auteurs utilisent des mesures d'entropie généralisées permettant de décomposer les inégalités en deux fractions, inégalités interindividuelles de revenu total calculé sur le cycle de vie d'une part, inégalités intra-individuelles au cours du cycle de vie d'autre part. Cette mesure des inégalités dépend d'un paramètre d'aversion à la pauvreté : lorsque celui-ci vaut 0 et 1 respectivement, on retrouve l'indice de Theil-L  $I_0$  et l'indice de Theil  $I_1$  respectivement. Le premier correspond à un degré plus grand d'aversion à la pauvreté, accordant une importance relativement plus grande aux revenus très faibles pour une période donnée. Les formules correspondant à ces deux indices sont les suivantes :

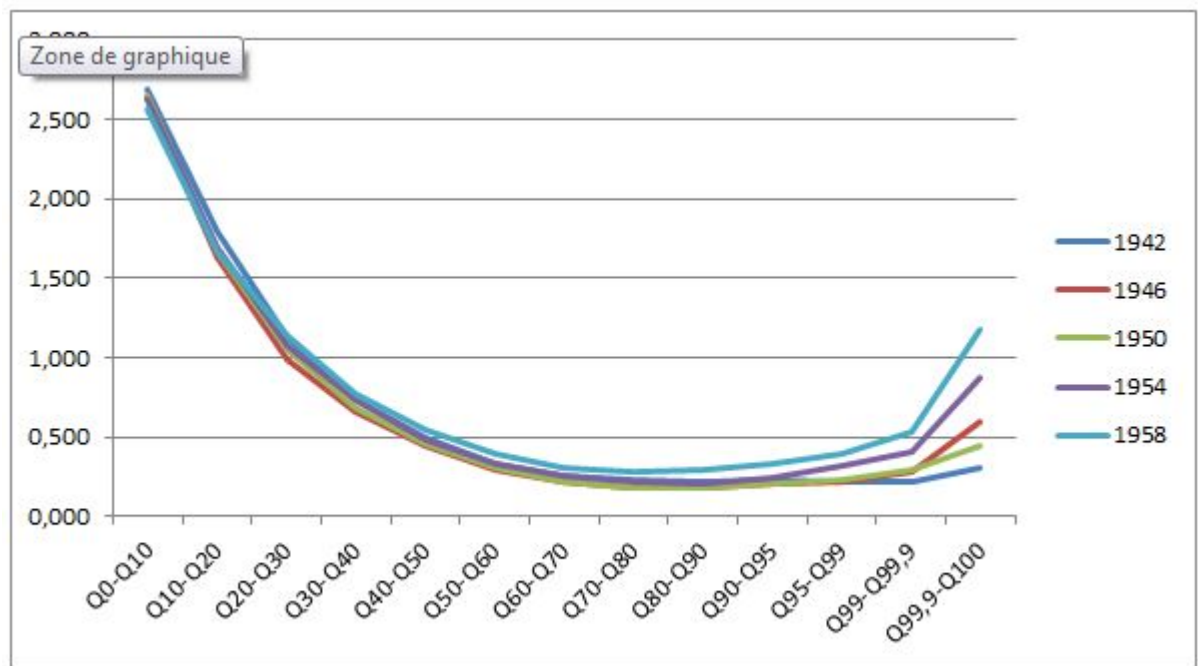
$$I_1 = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{\bar{y}} \log\left(\frac{y_i}{\bar{y}}\right)$$

$$I_0 = \frac{1}{n} \sum_{i=1}^n \log\left(\frac{\bar{y}}{y_i}\right)$$

avec  $y_i$  le revenu de l'individu à la période  $i$  et  $\bar{y}$  son revenu moyen sur le cycle de vie.  $I_0$  ne peut être calculé que dans le cas où les revenus sont non-nuls pour tous les individus à toutes les périodes. Cela n'est pas le cas pour nos trajectoires salariales. L'indice de Theil  $I_1$  au contraire peut être calculé même lorsque le revenu des individus est nul à certaines périodes, en ne prenant pas en compte ces périodes dans la somme (on a  $\lim_{y \rightarrow 0} y \log(y) = 0$ ). On calcule alors un indice de Theil intra-individuel pour chaque individu, sur la période 1976 à 2001 et pour les générations 1942 à 1958 (la génération 1958 a entre 18 et 43 ans et la génération 1942 entre 34 et 59 ans sur la période considérée). Cet indice individuel renseigne sur la volatilité des revenus sur le cycle de vie. Björklund et Palme considèrent quant à eux les revenus des individus sur une période de 18 années allant de 1974 à 1991, pour une catégorie d'individus " âgés " (ayant entre 33 et 47 ans en 1974) et une catégorie d'individus " jeunes " (ayant entre 18 et 32 ans en 1974). Pour chacun de ces deux groupes, ils examinent la corrélation entre l'indice de Theil intra-individuel (sur le cycle de vie) et le revenu total sur la période considérée. Celle-ci est significativement négative, signifiant que les carrières salariales sont plus instables dans le bas de la distribution. Lorsqu'ils effectuent la même analyse quartile par quartile, les auteurs trouvent une corrélation significativement négative pour le premier quartile, mais non significativement différente de zéro pour les trois quartiles de revenus (sur le cycle de vie) les plus élevés. Toutes générations confondues (1942 à 1958), on trouve pour les données EIC-DADS une corrélation entre indice de Theil intra-individuel et revenus salariaux sur le cycle de vie significativement négative (au seuil de 1 %), d'une valeur de -0.48. Lorsqu'on divise la population en quartiles, les résultats sont les suivants :

	P0-25	P25-50	P50-75	P75-100
Indice de Theil moyen	1.95	0.67	0.28	0.23
Corrélation avec le revenu total	-0.74 (***)	-0.47 (***)	-0.18 (***)	0.18 (***)

Et lorsqu'on examine les individus génération par génération, et en découpant la distribution à un niveau plus fin (particulièrement pour le haut de celle-ci), on peut figurer graphiquement les indices de Theil moyens par quantile :



On observe ainsi une baisse de l'instabilité des revenus salariaux jusqu'à un certain point, et une remontée assez prononcée dans l'extrémité haute de la distribution des revenus. Ceci pourrait être dû en particulier à la prise en compte des bonus et primes dans les salaires, bonus et primes qui sont variables au cours du temps et pouvant représenter des montants considérables pour les salariés les plus rémunérés.

## 2 Simulation sur le cycle de vie

It's like in chess : First, you  
strategically position your pieces  
and when the timing is right  
you...strike.

*David Levinson*

INDEPENDANCE DAY

---

Cet épigraphe est là pour rappeler que le projet TaxIPP-Life est encore très jeune. Dans une phase de développement, les résultats obtenus ont pour l'instant comme principal voire unique mérite d'être produits par le modèle. Il s'agit pour l'instant plus d'une réalisation technique et donc un peu frustrante. La phase de production de résultats ne devrait toutefois pas tarder.

Avant de commencer cette partie, certaines personnes doivent être remerciées. En effet, le projet TaxIPP-Life a bénéficié de l'aide de plusieurs autres projets. Le souci de partage des personnes concernées, leur réceptivité aux idées d'amélioration ainsi que leurs conseils et leurs temps ont été d'une aide extrêmement précieuse. TaxIPP-Life n'est pas encore un sujet abouti mais il serait très éloigné de ce qu'il est maintenant et auraient probablement des objectifs beaucoup plus restreint que ce qu'il a. Je remercie donc l'équipe d'OpenFisca<sup>4</sup> et en particulier M. Ben Jelloul, l'équipe de Liam2<sup>5</sup> en en particulier G. Dekkers et G. de Menten, l'équipe Destinie de l'Insee avec Aude Leduc et Anthony Marino et l'équipe de PensIPP<sup>6</sup> et en particulier D. Blanchet et S.Rabaté.

### 2.1 Quelques mots de technique

Dans cette courte section, seront décrits les choix techniques faits pour réaliser la simulation. Nous verrons que la question de la performance ne doit pas être négligée. Appliquer la législation socio-fiscale d'une année prend plus de 15 minutes lorsqu'elle est codée en SAS par exemple. Dans un calcul sur l'ensemble du cycle de vie, cette législation doit être simulée disons cent fois, l'optimisation du temps de calcul peut donc avoir du bon pour ne pas prendre 25 heures de calcul. Cela d'autant plus que la volonté de travailler sur un gros échantillon peut augmenter considérablement le temps de calcul, et que l'on a l'idée un jour d'introduire de la décision comportementale des agents en fonction de leur espérance de gain actualisée, et donc de simuler la législation encore bien plus de cent fois.

#### 2.1.1 La base de données initiales

Comme on peut l'imaginer à la lecture de la première partie de ce document c'est l'enquête Patrimoine qui va servir de base d'information à notre modèle. La raison en est la connaissance

---

4. <http://www.openfisca.fr/>

5. <http://liam2.plan.be/>

6. <http://www.ipp.eu/fr/outils/pensipp-simulation/>



des trajectoires professionnelle passée. La connaissance du patrimoine est aussi , bien sûr, une des forces de l'enquête. On aurait pu utiliser d'autres sources. Par exemple l'enquête Santé et Itinéraire Professionnel, qui comme son nom l'indique, est bien moins précise sur le patrimoine mais bien plus sur les conditions de santé. Pour la simulation du futur uniquement, on peut aussi envisager l'utilisation d'enquêtes plus générales et n'ayant pas de volet trajectoire. Ainsi, l'enquête Budget des Familles est par exemple attirante car la consommation est alors connue, au moins pour une année. On peut aussi penser à l'enquête Logement. L'idée est à étudier mais on peut essayer d'étendre le matching statistique réalisé avec deux enquêtes à plusieurs en constituant ce faisant une sur-enquête couvrant la majorité des champs micro-économiques.

Dans la suite, nous ne parlerons que de l'enquête Patrimoine 2009-2010. L'échantillon contient 29 951 individus répartis en 12788 ménages<sup>7</sup>. De plus, 14 954 déclarations fiscales sont imputées (pour l'instant sommairement).

### 2.1.2 La fermeture de l'échantillon

Une première limite de l'enquête Patrimoine pour notre sujet est qu'elle ne donne pas de lien entre les ménages. Si les relations sont connues au sein d'un ménage, on veut aussi connaître les relations et particulièrement les liens de filiation entre les individus qui ne vivent pas dans le même ménage. Cela peut permettre d'imputer ensuite les successions, les pensions alimentaires l'aide pendant les études mais aussi les différentes formes d'aide vers les ascendants, en particulier en cas de dépendance. On peut aussi imaginer d'imputer un recours aux grands parents pour les gardes d'enfants.

Dans l'enquête patrimoine, les parents ayant des enfants vivant à l'extérieur du domicile se voient poser quelques questions à leur sujet (date de naissance, nombre d'enfant, statut sur le marché de l'emploi, diplôme, etc.). Nous créons ainsi des individus factices représentant ces enfants hors domicile. Puis par un procédé de matching, nous cherchons des individus correspondants, déclarant un ou deux parents en vie, qui correspondent le plus possible à ces statistiques. On attribue alors les parents enquêtés à l'individu fictif comme étant ses parents

Nous n'utilisons pas les informations ascendantes, celles que les enfants déclarent à propos de leur parents. Elles sont assez pauvres et concernent pratiquement exclusivement l'enfance de l'enfant or il se peut que la situation des parents ait changé entre l'enfance de l'enfant et le moment de l'enquête. On pourrait aussi utiliser les informations sur le décès ou non des grands-parents. Enfin, notons que nous n'utilisons que les données de l'enquête, on n'utilise pas d'information annexe par exemple sur la corrélation entre les revenus des parents et des enfants. Pour l'instant, on fait donc l'hypothèse que les variables utilisées sont suffisantes pour que les distributions croisées soient respectées.

Nous donnons ici une précision sur la méthode de calcul pour souligner au lecteur une petite

---

7. Les ménages interrogés aux Antilles, ont un questionnaire légèrement différent. Par souci de simplicité et pour travailler avec un échantillon uniforme, ils sont pour l'instant exclus du champ.

limite actuelle de cette étape. Lorsque les parents sont interrogés sur leurs enfants vivant hors du domicile, on leur demande de préciser s'il s'agit de l'enfant du couple ou de l'un d'eux seulement. S'il s'agit de l'enfant du couple, alors on cherche des enfants dont les deux parents sont vivants. Il reste ensuite de tels enfants avec deux parents vivants à apparier ce qui est logique car les parents peuvent être vivants et ne plus vivre ensemble. On cherche pour eux, comme pour les enfants n'ayant qu'un parent à trouver, un père puis une mère. Si on ne cherche qu'un parent, il n'y a pas de problème. En revanche, un point à noter est que lorsque l'on cherche deux parents, on les cherche indépendamment. Par exemple, si on a deux enfants identiques, deux pères potentiels, l'un de 40 et l'autre de 60 ans et deux mères potentielles, l'une de 40 ans et l'autre de 60 ans, rien ne contrôle que l'on a plus de chance d'avoir un couple dont les deux membres ont 40 ans et un dont les deux couples ont 60 ans que deux couples avec 20 ans d'écart d'âge. On peut donc ne pas reproduire la réalité statistique en créant une distribution jointe des parents non réaliste. Du moins c'est le cas, si on pense que deux couples de parents peuvent avoir des enfants similaires même s'ils sont différents au départ. Une solution serait, une fois la mère trouvée, de chercher un père avec un âge assez proche, une autre solution serait de réaliser un mariage fictif des parents ayant des enfants similaires puis de tirer un couple de parents pour les enfants.

Enfin, précisons que le matching est fait à partir d'une méthode de score (voir Annexe pour plus d'information sur les méthodes de matching)

### 2.1.3 Étendre l'échantillon

Travailler avec des pondérations n'est en général pas un problème dans le cadre de statistiques statiques. Cependant, dans le contexte particulier de la microsimulation dynamique où des liens sont établis entre ménages cela pose problème. On ne peut en effet pas associer 50 enfants à 3 000 mères ou unir un homme en représentant 1000 avec une femme représentant 2000 femmes de la population française. On se demande par exemple, comment on pondérerait leurs enfants. Pour solutionner le problème, le choix a été fait de dupliquer les individus autant de fois que nécessaire pour ainsi avoir une pondération uniforme. Dans l'enquête Patrimoine 2010, qui contient un sur-échantillonnage des hauts patrimoines, la plus petite pondération est de 6 et cela devrait être la pondération uniforme. Cela nous mène en théorie à une base de plus de 10 millions de lignes dans la base pour représenter la population française.

Le modèle DESTINIE tire ensuite un sous-échantillon à partir de cette population étendue. TaxIPP-Life, quant à lui, essaie de tourner sur l'ensemble de cet échantillon. Cela a aussi l'avantage qu'un même individu au départ se verra associé à plusieurs carrières. On conserve ainsi une certaine variabilité dans le futur de chaque individu de la base initiale ce qui ne peut être que bon pour la précision statistique. Comme nous le verrons ci-dessous, travailler sur une base étendue est un vrai challenge. Pour l'instant, nous travaillons en général sur l'échantillon non étendu, sans tenir compte des pondérations dans les différentes étapes. Une alternative acceptable en

termes de temps de calcul et qui permet de tester le modèle est d'utiliser comme pondération uniforme 200, ce qui fait une base de 300 000 individus. Les ménages avec une pondération initiale inférieure à 200 se voient attribuer un poids de 200, ce qui biaise donc en théorie pour l'instant un peu les résultats.

#### 2.1.4 Choix de langage de programmation

Toutes les étapes initiales, y compris l'extension de l'échantillon décrit ci-dessus, sont réalisées en R. Les étapes de simulation, vieillissement de la population et calcul de la législation sont réalisées en Python. On utilise ainsi les forces de chaque programme<sup>8</sup>. L'utilisation de Python permet une écriture adaptée à la microsimulation et est relativement incomparable avec les logiciels statistiques en termes de performances dont on a montré qu'elles étaient importantes pour TaxIPP-Life. L'étape de calcul de la législation qui prend 15 minutes en SAS est effectué dans un temps de l'ordre de grandeur de dix secondes. La possibilité de travailler avec un gros échantillon ou avec des équations comportementales devient envisageable.

### 2.2 Les étapes de la simulation dynamique

Avant d'entamer cette section, il faut préciser que TaxIPP-Life est encore un projet jeune en développement. Les étapes présentées ici peuvent toutes être améliorées, certaines sont tellement rudimentaires qu'elles n'ont d'autre mérite que d'exister. La description de ces étapes permet toutefois de donner une première idée des fonctionnalités et possibilités de TaxIPP-Life. L'écriture sous forme de modules indépendants permet à tout moment d'améliorer chacune de ces étapes.

#### 2.2.1 Démographie

L'âge est bien sûr incrémenté à chaque période.

**Naissance** Toutes les femmes en couple âgées de 16 à 50 ans se voient imputer un enfant<sup>9</sup>. Un certain nombre de ces femmes est sélectionné en s'assurant de vérifier les projections de naissance en fonction de l'âge de la mère. Le sexe de l'enfant est choisi aléatoirement.

Le fait de ne pas avoir de probabilité différente pour les femmes selon leurs caractéristiques autre que leur âge est problématique. Le nombre d'enfants ainsi que le niveau d'études devrait intervenir<sup>10</sup>.

---

8. Parce qu'il s'agit d'un matching, l'étape de création des liens entre parents et enfants vivant hors domicile est, elle aussi, réalisée en Python.

9. On élimine les femmes qui ont 8 enfants ou plus ou dont le conjoint est dans cette situation. On pourrait toutefois supprimer sans difficulté cette condition.

10. On ne peut pas pour autant dire que le tirage obtenu est complètement indépendant du niveau d'études et du nombre d'enfants dans la mesure où ceux-ci interviennent dans les équations de mise en couple et de divorce, pré-requis pour se voir imputer des enfants dans TaxIPP-Life.

**Décès** La probabilité de décès ne dépend que de l'âge et du sexe. Le nombre de décès prédit par les projections démographiques de l'Insee est ainsi reproduit. Si une personne décédée vit seule avec des enfants, ces enfants sont attribués à l'autre parent si celui-ci est encore en vie. Si ce n'est pas le cas, ils sont affectés à un ménage spécial sans adulte avec tous les autres enfants dans ce cas.

**Union** Pour l'instant, dans le modèle, l'union entre deux personnes correspond à la fois à un emménagement et à une union légale. Sans qu'il n'y ait de contre-indication technique à le faire, il n'y a donc pas de création de concubinage. Il en existe toutefois dans la base initiale. Reprenant les équations utilisée par DESTINIE à partir du travail de Duée(??) on impute une probabilité d'avoir se mettre en couple. Cela est fait d'une part pour les premières unions et d'autre part pour les personnes ayant déjà été en couple séparément selon le sexe, en fonction de l'âge depuis la fin des études, du fait d'avoir un enfant, et de la durée depuis la précédente séparation le cas échéant. En utilisant une méthode d'alignement de Liam2 (voir ...), un tiers de ces personnes est sélectionné<sup>11</sup>. Ensuite, un matching est effectué pour associer les individus en fonction de leur âge, de leur différence d'âge et de leur niveau de diplôme comparé. Encore une fois, ce matching, crucial pour étudier bon nombre de questions de redistribution, par exemple, l'impact de l'endogamie, mériterait d'être affiné. En cas d'union, les personnes rattachées aux nouveaux conjoints (enfants le plus souvent) emménage avec eux. Idem pour les déclarations fiscales.

**Séparation** C'est plus crédible cette fois : les séparations se traduisent à la fois par un changement de logement d'un des deux conjoints et par une rupture du contrat (et donc le passage à des déclarations fiscales séparées). La probabilité de rupture dépend du nombre d'enfant du couple, de son ancienneté et de la différence d'âge entre ses membres. Ceci pourrait être régi par des probabilités mais pour l'instant c'est l'homme qui déménage (sauf lorsque le couple habitait chez ses parents) et la femme qui a une nouvelle déclaration<sup>12</sup>. Les enfants et personnes rattachées ne changent de logement que s'ils sont liés à la personne qui part sans être liés à la personne qui reste, que ce soit pour le logement ou pour la déclaration fiscale.

### 2.2.2 Logement

Le travail sur le logement est extrêmement rudimentaire. Les déménagement ne se font que lors des unions et séparations et lorsqu'une personne de plus de 24 ans n'est ni personne de référence ni en couple (i.e. vit avec ses parents). Le travail sur les loyers n'a pas encore été fait mais ne devrait pas poser de difficulté majeure.

On pourrait à terme utiliser réellement un parc de logement dont on simulerait la variation chaque année. Lors des déménagements, les ménages choisiraient l'un des logements en fonction de la taille du ménage et du logement ainsi qu'en fonction des revenus et du loyer. On simulerait

---

11. Un calage plus pertinent ne serait pas une mauvaise chose.

12. Dans le cas de couple homosexuel, le choix est fait aléatoirement

ainsi des déménagements en fonction d'un surpeuplement ou sous-peuplement ainsi qu'en cas d'évolution dans les revenus. L'idée serait de reproduire des tensions sur le marché du logement.

On impute aussi initialement pour chaque logement un propriétaire. Du travail est encore nécessaire à cette étape mais indubitablement savoir identifier les propriétaires de leur résidence, les locataires sera un plus pour le modèle.

### 2.2.3 Marché du travail

La question de la situation sur le marché de l'emploi est évidemment un élément crucial pour TaxIPP-Life, qui veut étudier la redistribution. C'est à ce niveau que l'investissement le plus important devra être fait. Les équations de transition devraient être estimées le plus précisément possible en tenant compte d'un nombre important de paramètres afin de reproduire la réalité du marché du travail. On est pour l'instant loin d'un niveau satisfaisant mais il sera, là encore, facile, au moins en termes de programmation d'améliorer cette partie du modèle.

**Éducation** L'entrée sur le marché du travail correspond à la fin des études. L'âge de fin d'études est déterminé à partir de l'âge de fin d'études des parents (au moment de la naissance de l'enfant donc). On peut penser à des modèles plus compliqués de décision (voir 2.3.3). Pour l'instant, on impute pour les individus qui ne sont pas encore sortis du système scolaire un âge de fin d'études qui est la moyenne de celui de ses parents auxquels on ajoute un élément correcteur en fonction de leur date de naissance pour tenir compte de l'allongement moyen des études.

Lorsque l'individu est encore en étude, son niveau scolaire est entièrement déterminé par son âge. Ce niveau d'étude servira ensuite à imputer la valeur du transfert en nature associée à l'éducation.

**Statut** Dans l'idéal, on détermine l'état à la date  $T+1$  en fonction de l'état à la date  $T$  mais aussi des états précédents pour être plus précis. Pour l'instant deux méthodes simplistes ont été implémentées. La première calcule un score en fonction de l'âge puis tire un nombre de transition en suivant des proportions en input du modèle. La seconde sélectionne aléatoirement 2,5 % des individus actifs et inactifs pour les faire changer de statut. Au passage à l'activité, la catégorie d'emploi est tirée au sort.

**Type d'emploi et d'employeur** Comme pour le logement, on crée un fichier d'entreprise. Tout est à faire mais on a avec l'enquête Patrimoine une idée de la taille de l'entreprise des salariés. Une des entreprises peut regrouper les indépendants et libéraux. Cet élément n'est pas prioritaire mais l'idée est de pouvoir simuler un marché du travail par secteur ou bien d'avoir un modèle de travailleurs indépendants ou encore de se focaliser sur un secteur d'activité particulier, la médecine par exemple.

**Salaire et revenus** Les salaires sont pour l'instant imputés à partir d'une équation de Mincer relativement simple. Si la personne a un revenu d'activité initialement, on détermine une productivité individuelle comme étant le rapport entre son salaire et le salaire de l'équation de

Mincer. Pour les autres, la productivité est tirée aléatoirement. Par la suite, l'équation de Mincer sera toujours multipliée par la productivité pour déterminer les salaires.

**Retraite** Le passage à la retraite est pour l'instant entièrement déterminé par l'âge.

#### **2.2.4 Épargne et consommation**

Les deux thèmes d'épargne et de consommation doivent toujours être pensés ensemble. En effet, à chaque période, on sait que le revenu est consommé ou épargné. Si l'un de deux éléments est déterminé, l'autre s'en déduit. Un taux d'épargne est donc simulé. Il faut ensuite déterminer la structure de consommation ainsi que la structure d'épargne. L'achat immobilier demandera une imputation particulière.

On se servira de l'enquête Budget des Familles, pour imputer une structure de consommation en catégorie permettant de calculer les taxes sur la consommation.

A partir du nouveau stock d'épargne de la période  $T$ , on détermine des revenus du patrimoine à la période  $T+1$ . Lors du décès d'un individu, on répartit son capital parmi ses héritiers connus dans le modèle (conjoint, enfants et parents). Si l'individu n'est lié à personne au moment de son décès, on considère pour l'instant le capital perdu. Dans un avenir proche, les droits de succession seront aussi calculés.

#### **2.2.5 Calcul de la législation**

. En fin de simulation, la législation est calculée. Cela inclut les cotisations salariales, l'impôt sur le revenu, l'impôt sur la fortune, le bouclier fiscal et les prestations. Les prestations associées au handicap sont déjà simulées mais cette information n'est pas encore dans la projection de la population. Les taxes (habitation et foncière) sont encore à travailler. Tout le volet santé est aussi à travailler.

Enfin, les retraites et le chômage devrait être imputés d'ici peu.

### **2.3 Projets futurs**

On a déjà donné beaucoup de pistes d'amélioration dans la section précédente. Ici, on présente les éléments qui pourront agrémente le modèle à l'avenir.

#### **2.3.1 Simulation par mois**

Les modèles de microsimulation travaillent usuellement sur une base annuelle. Pourtant la plupart des minima ainsi que les retraites sont calculées par trimestre, les prestations familiales et logement ainsi que les cotisations et contributions prélevées à la source sont calculées mois par mois. La simulation dynamique étant rapide, on peut modifier la période pour que le pas devienne le trimestre ou le mois. L'étape de mise en couple qui est la plus longue peut éventuellement

n'être simulée qu'une fois par an pour gagner encore du temps de calcul<sup>13</sup>. L'application de la législation pourrait se faire d'autant plus précisément.

### 2.3.2 Transferts intergénérationnels

Dans un premier temps, les calculs sont effectués avec une législation figée, celle de 2009. Les revenus et la population future sont aussi simulés en gardant les paramètres de 2009. Par exemple, il n'y a pas de croissance simulée. Les revenus passés sont eux revalorisés en fonction de la hausse des revenus pour se voir appliquer la législation de 2009. Plus généralement toutes les équations pour obtenir des informations non contenues dans la base de départ sont simulées à partir des estimations les plus récentes. Pourtant, les comportements ont changé (la consommation de téléphones portables était réduite dans les années 80 par exemple) et si l'on voulait avoir une représentation fidèle de la consommation passée des individus par exemple, il faudrait en tenir compte.

Bien évidemment, à terme, TaxIPP-Life simulera l'ensemble des législations passées et simulera la population en fonction de tendances observées ou supposées. Les imputations dépendront de la période simulée. Ceci permettra de reproduire fidèlement la population et de pouvoir procéder à une analyse d'équité intergénérationnelle. Pour l'instant, on étudie finalement une population de 2009 évoluant et ayant évolué dans les conditions de 2009 hormis pour les trajectoires salariales et matrimoniales avant 2009.

### 2.3.3 Comportements

Une dimension supplémentaire peut être apportée au modèle en introduisant des équations comportementales. Au lieu de simuler des transitions en fonction d'équations estimées, on peut créer un modèle d'agent dans lequel les individus prennent leur décision en fonction d'anticipation. Il y a deux décisions importantes dans le cycle de vie dans lesquelles ces modèles ont leur place : le choix d'éducation et la retraite.

Le choix d'éducation est le plus compliqué des deux car les anticipations rationnelles portent sur une durée plus grande et la diversité des trajectoires envisagée peut être très importante. L'idée serait, pour un individu devant décider à la période  $T$  d'entrer sur le marché du travail ou continuer ses études, de générer une évaluation de ses revenus dans les deux options et de lui faire choisir celle qui optimise son utilité (actualisée). Ainsi, on pourrait mesurer l'influence d'une hausse de l'impôt sur le revenu, des minima sociaux ou encore des retraites sur les choix d'éducation. Bien sûr, la carrière n'est pas entièrement déterminée par le choix d'éducation, on pourra, dans les deux options, simuler plusieurs centaines d'avenirs possibles et estimer ainsi l'espérance de l'utilité de chaque option.

---

13. Il s'agit là d'un abus de langage, car si on travaillait par trimestre par exemple, on aura quatre fois un quart des mariages annuels à simuler. Comme l'algorithme de matching est non linéaire, il est plus rapide de faire tourner quatre fois sur le quart de l'échantillon qu'une fois sur l'échantillon global. La contrepartie en est que le matching obtenu est théoriquement de moins bonne qualité puisque chaque individu n'a qu'un choix limité d'époux. Cette partie demanderait à être expertisée.

### 2.3.4 Autres enquêtes

Comme on l'a dit en introduction de cette partie, le modèle pourrait aussi gagner à utiliser d'autres enquêtes. En particulier l'utilisation d'enquêtes portant sur la santé pourrait ajouter cette dimension au modèle.

## 3 Résultats préliminaires : simulation des allocations familiales (...et de ??)

### Annexe

#### Méthode de matching

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Salarié du public à temps complet (1)	0	8600	2900	12400	5300	15120	11900	19260	7100	23300
Salarié du public à temps partiel (2)	8600	0	5700	3800	3300	6520	3300	10660	1500	14700
Salarié du privé à temps complet (3)	2900	5700	0	9500	2400	12220	9000	16360	4200	20400
Salarié du privé à temps partiel (4)	12400	3800	9500	0	7100	2720	500	6860	5300	10900
A son compte (5)	5300	3300	2400	7100	0	9820	6600	13960	1800	18000
Chômage (6)	15120	6520	12220	2720	9820	0	3220	4140	8020	8180
Succession de périodes d'emploi et de chômage (7)	11900	3300	9000	500	6600	3220	0	7360	4800	11400
Reprise d'études ou formation (8)	19260	10660	16360	6860	13960	4140	7360	0	12160	4040
Préretraite (9)	7100	1500	4200	5300	1800	8020	4800	12160	0	16200
Inactif et autres (10)	23300	14700	20400	10900	18000	8180	11400	4040	16200	0

Tableau résumant la qualité du matching selon les poids relatifs utilisés pour les distances ?

Tableau décrivant les salaires imputés par rapport aux années adjacentes ?

## Références

[Bemis and Morgan, 1975] Bemis, J. and Morgan, J. N. (1975). Time period, unit of analysis, and income concept in the analysis of income distribution. In *The Personal Distribution of*



- Income and Wealth*, NBER Chapters, pages 209–224. National Bureau of Economic Research, Inc.
- [Bengtsson et al., 2012] Bengtsson, N., Holmlund, B., and Waldenström, D. (2012). Lifetime versus annual tax progressivity : Sweden, 1968-2009. Working Paper Series, Center for Fiscal Studies 2012 :8, Uppsala University, Department of Economics.
- [Bjorklund, 1993] Bjorklund, A. (1993). A comparison between actual distributions of annual and lifetime income : Sweden 1951-89. *Review of Income and Wealth*, 39(4) :377–86.
- [Björklund and Palme, 1997] Björklund, A. and Palme, M. (1997). *Income Redistribution Within the Life Cycle Versus Between Individuals : Empirical Evidence Using Swedish Panel Data*. Working paper series in economics and finance. Stockholm School of Economics, the Economic Research Inst.
- [Bourguignon, 1998] Bourguignon, F. (1998). *Fiscalité et redistribution, La France dans une perspective internationale*. Rapport du Conseil d'Analyse Économique.
- [Duval et al., 2011] Duval, J., Eidelman, A., Langumier, F., and Lejbowicz, T. (2011). La redistribution : état des lieux en 2011. *France, Portrait social*.
- [Esping-Andersen, 1990] Esping-Andersen, G. (1990). *The three worlds of welfare capitalism*. PRINCETON University Press.
- [Falkingham et al., 1996] Falkingham, J., Harding, A., of Canberra. National Centre for Social, U., and Modelling, E. (1996). *Poverty Alleviation Versus Social Insurance Systems : A Comparison of Lifetime Redistribution*. Discussion paper. University of Canberra.
- [Fitzgerald and Maloney, 1990] Fitzgerald, J. and Maloney, T. (1990). The impact of federal income taxes and cash transfers on the distribution of lifetime household income : 1969-1981. *Public Finance Quarterly*, 18(2) :182–197.
- [Fragonard et al., 2013] Fragonard, B., Hot, E. L., Leprince, F., and Bonnevide, P. (2013). Les aides aux familles. Technical report, Haut Conseil de la Famille.
- [Hoynes and Luttmer, 2011] Hoynes, H. W. and Luttmer, E. F. (2011). The insurance value of state tax-and-transfer programs. *Journal of Public Economics*, 95(11-12) :1466 – 1484. Special Issue : International Seminar for Public Economics on Normative Tax Theory.
- [Hussenius and Selen, 1994] Hussenius, J. and Selen, J. (1994). *Skatter och socialförsäkringar över livscykeln - en simuleringsmodell*. departementsserien. Finansdepartementet.
- [Lindert, 2004] Lindert, P. (2004). *Growing Public : Volume 1, The Story : Social Spending and Economic Growth since the Eighteenth Century*. Cambridge University Press.
- [Pettersson and Pettersson, 2007] Pettersson, T. and Pettersson, T. (2007). *ModeModel Our Future - Population Ageing, Social Security and Taxation*, chapter Lifetime Redistribution Through Taxes, Transfers and Non-Cash Benefits. Elsevier.
- [Shorrocks, 1978] Shorrocks, A. (1978). Income inequality and income mobility. *Journal of Economic Theory*, 19(2) :376–393.

- [Slemrod, 1992] Slemrod, J. (1992). Taxation and inequality : A time-exposure perspective. NBER Working Papers 3999, National Bureau of Economic Research, Inc.
- [Varian, 1980] Varian, H. R. (1980). Redistributive taxation as social insurance. *Journal of Public Economics*, 14(1) :49–68.