



Institut des
Politiques Publiques

GUIDE METHODOLOGIQUE IPP – NOVEMBRE 2019

Le modèle de microsimulation TAXIPP – Version 1.1

Mahdi BEN JELLOUL
Antoine BOZIO
Thomas DOUENNE
Brice FABRE
Claire LEROY





L'Institut des politiques publiques (IPP) est développé dans le cadre d'un partenariat scientifique entre PSE-Ecole d'économie de Paris (PSE) et le Centre de Recherche en Economie et Statistique (CREST). L'IPP vise à promouvoir l'analyse et l'évaluation quantitatives des politiques publiques en s'appuyant sur les méthodes les plus récentes de la recherche en économie.

www.ipp.eu





GUIDE METHODOLOGIQUE – NOVEMBRE 2019

Le modèle de micro-simulation TAXIPP – Version 1.1

Mahdi BEN JELLOUL
Antoine BOZIO
Thomas DOUENNE
Brice FABRE
Claire LEROY

SOMMAIRE

| | | |
|----------|--|-----------|
| 1 | Présentation générale | 3 |
| 2 | Les données sources | 7 |
| 2.1 | Les bases de données mobilisées | 7 |
| 2.1.1 | Les données des fichiers de l'impôt sur le revenu (FELIN) . . . | 8 |
| 2.1.2 | Le fichier panélisé des déclarations ISF/IFI | 8 |
| 2.1.3 | Les données des fichiers de la taxe d'habitation et des propriétés bâties (FIDELI) | 9 |
| 2.1.4 | L'enquête "Revenus Fiscaux et Sociaux" (ERFS) | 9 |
| 2.1.5 | Les données de l'enquête "Budget de Famille" (BDF) | 9 |
| 2.2 | Fusion des données FELIN et ERFS | 10 |
| 2.2.1 | Harmonisation préalable des deux bases | 12 |
| 2.2.2 | Appariement statistique des foyers fiscaux | 13 |
| 2.2.3 | Création de tables individus, foyers fiscaux, familles et ménages | 16 |
| 2.2.4 | Viellissement de la base FELIN-ERFS | 20 |
| 3 | Programmes de simulation | 23 |
| 3.1 | Connection avec OpenFisca | 23 |
| 3.1.1 | La gestion des périodes | 24 |
| 3.1.2 | La reconstitution des revenus bruts | 25 |
| 3.2 | Simulation du système socio-fiscal | 26 |
| 3.2.1 | L'impôt sur le revenu | 27 |
| 3.2.2 | Les cotisations sociales | 28 |
| 3.2.3 | Les prélèvements sociaux | 29 |
| 3.2.4 | Les prestations familiales | 29 |
| 3.2.5 | Les aides au logement | 30 |
| 3.2.6 | Les minima sociaux | 31 |
| 3.2.7 | L'imposition du patrimoine | 32 |
| 3.2.8 | La taxe d'habitation | 34 |
| 3.2.9 | La fiscalité indirecte | 35 |
| 3.2.10 | Assemblage des différents modules | 37 |

| | |
|---|-----------|
| 4 Evaluations de réformes socio-fiscales | 39 |
| 4.1 La définition d'une réforme | 39 |
| 4.2 Les résultats produits | 41 |
| Conclusion générale | 43 |
| Références | 45 |

CHAPITRE 1

PRÉSENTATION GÉNÉRALE

Le modèle TAXIPP est un modèle de micro-simulation du système fiscal et social français développé par l'Institut des politiques publiques (IPP), un partenariat scientifique conjointement porté par PSE-École d'Économie de Paris et le Centre de recherche en économie et statistique (CREST).

Un document méthodologique. Ce guide méthodologique présente la version 1.1 du modèle et fait suite aux guides des versions antérieures TAXIPP 0.0 à TAXIPP 1.0¹. Il décrit le fonctionnement du modèle dans le détail, depuis les données sources utilisées aux programmes de simulation appliqués.

Les nouveautés de la version 1.1. Par rapport à la version 1.0, la version 1.1 simule l'impôt de solidarité sur la fortune (ISF) et l'impôt sur la fortune immobilière (IFI) à partir des données administratives associées à ces deux impôts. Cette nouvelle version s'affranchit donc des méthodes de déduction des patrimoines à partir de leurs revenus, employées précédemment. Cette version 1.1 fait suite au récent passage à la version 1.0, qui a représenté un changement important tant dans les sources que dans l'architecture. Les données utilisées sont issues d'un appariement statistique entre les données fiscales des déclarations de revenus (échan-

1. Voir Ben Jelloul et al. (2019b) pour la méthodologie de la version 1.1.

tillon FELIN) et de l'enquête ERFS. L'idée centrale est d'utiliser les données administratives autant que possible lorsque celles-ci sont disponibles. Les informations de sources administratives² sont généralement plus fiables que les données d'enquête, et leur échantillonnage est également plus fin, notamment au niveau des hauts revenus dans notre cas. En conséquence, cette version du modèle n'utilise que de façon marginale le calage sur les grands agrégats de la comptabilité nationale (contrairement aux versions précédentes), et le modèle fonctionne au sein du Centre d'accès sécurisé distant (CASD), les données sources étant accessibles grâce à cette technologie. Enfin, le modèle TAXIPP 1.1 repose sur l'utilisation du calculateur opensource OpenFisca co-développé par l'IPP, en partenariat avec une communauté diverse (<https://fr.openfisca.org/>). TAXIPP 1.1 est donc codé en langage Python, comme OpenFisca, contrairement aux versions précédentes codées en Stata. Le calculateur OpenFisca vise à simuler l'ensemble de la législation socio-fiscale et peut être utilisé dans des cas divers (simulations avec données, API web). Il s'appuie notamment sur les *Barèmes IPP*, disponibles en ligne (<https://www.ipp.eu/outils/baremes-ipp/>), qui recensent l'historique des barèmes de la législation socio-fiscale française.

Un modèle statique, avec des réponses comportementales limitées. Le modèle TAXIPP 1.1 est un modèle « statique », dans le sens où il ne modélise pas les réponses comportementales, excepté pour la fiscalité indirecte. Autrement dit, l'estimation des effets redistributifs d'une réforme socio-fiscale repose sur l'hypothèse que les ménages ne modifient pas leurs comportements du fait de la réforme. Néanmoins,

2. Les données fiscales n'ont pas été accessibles hors de l'administration fiscale pendant très longtemps. Le cadre légal de leur accès était défini par l'article L135 du Livre de procédures fiscales qui fixe une interdiction de principe à leur diffusion, et une liste de dérogations précisées au L135-D. La loi 2013-660 du 22 juillet 2013 relative à l'enseignement supérieur et à la recherche (JO 23 juillet 2013) a modifié cet état de fait par son article 104 qui autorise un accès à ces données pour des fins de recherche scientifique. Le décret 2014-950 du 21 août 2014 précise les conditions d'application : passage devant le comité du secret statistique qui évalue la pertinence de la demande, validation par la DGFIP, signature de la part des chercheurs d'engagements de confidentialité, puis accès via le Centre d'accès sécurisé distant (CASD) aux données (<https://casd.eu/>). Voir Bozio et Geoffard (2017)

pour la fiscalité indirecte, nous modélisons les réactions des ménages en termes de consommation du fait de variations de prix. Les versions ultérieures du modèle proposeront des modules de réponses comportementales plus complets.

Bach et al. (2019b) constituent une exception, dans la mesure où ils incorporent, dans le cadre d'un exercice spécifique, des réponses comportementales à la fiscalité des dividendes dans l'évaluation de l'impact budgétaire de deux réformes majeures de la fiscalité du capital. Les réponses comportementales incorporées sont celles évaluées par Bach et al. (2019a), qui utilisent également le modèle TAXIPP pour caractériser les réformes évaluées.

Les travaux issus de la version 1.1. Le modèle TAXIPP 1.1 a été utilisé à l'IPP dans le cadre de plusieurs travaux d'évaluation de politiques publiques. Fabre et al. (2019) utilisent ce modèle pour évaluer les impacts redistributifs des mesures socio-fiscales du projet de loi de finances (PLF) et du projet de loi de financement de la sécurité sociale (PLFSS) de 2020, ainsi que des mesures socio-fiscales entrées en vigueur de 2018 à 2019. Bach et al. (2019a) et Bach et al. (2019b) évaluent l'impact des récentes réformes de la fiscalité des dividendes sur les flux de revenus du capital. Ils utilisent TAXIPP 1.1 pour caractériser la nature de ces réformes.

CHAPITRE 2

LES DONNÉES SOURCES

Le modèle TAXIPP 1.1 repose sur l'utilisation de sources de données multiples, dont certaines sont appariées statistiquement afin de pouvoir simuler le système socio-fiscal sur une base de données unique. Un appariement statistique entre deux bases de données consiste à définir, pour chaque individu de la première base (la base « receveuse ») son « plus proche voisin » dans la deuxième base (la base « donneuse »). Il s'agit donc de créer des individus fictifs, dans le but de reconstituer les distributions jointes des variables présentes dans chacune des sources.

Cette partie décrit la création des bases de données individuelles fictives utilisées par le modèle pour simuler le système socio-fiscal français. La section 2.1 présente les sources de données mobilisées. La section 2.2 décrit en détail la méthode d'appariement de ces différentes sources.

2.1 Les bases de données mobilisées

L'une des innovations majeures du modèle TAXIPP depuis sa version 1.0 réside dans l'utilisation de données administratives. La plupart des modèles de microsimulation français reposent sur des données d'enquête mais leur utilisation soulève plusieurs problèmes, notamment concernant la taille limitée des échantillons et le

manque de fiabilité des informations auto-déclarées dans certains cas. L'utilisation de données administratives permet de dépasser ces limites, en se basant sur un nombre plus important de ménages, et sur des informations fiables utilisées par les administrations pour le calcul des dispositifs socio-fiscaux.

2.1.1 Les données des fichiers de l'impôt sur le revenu (FELIN)

Le Fichier Échantillonné de l'Impôt sur le Revenu (FELIN) est un échantillon de 500 000 déclarations fiscales représentatif des foyers fiscaux français. Les données sont publiées par la Direction Générale des Finances Publiques (DGFIP) chaque année, depuis 1994. Dans TAXIPP 1.1, nous utilisons le millésime 2014. Ces données permettent de connaître dans le détail l'ensemble de la déclaration fiscale d'un foyer. Par ailleurs, l'échantillonnage est exhaustif parmi les 1 % des foyers aux revenus imposables les plus élevés.

2.1.2 Le fichier panéalisé des déclarations ISF/IFI

Ce fichier recense sur la période 2006-2018 l'exhaustivité des déclarations à l'impôt de solidarité sur la fortune (ISF) et à l'impôt sur la fortune immobilière (IFI). Pour chaque déclaration sont renseignées l'ensemble des cases du formulaire rempli par les déclarants. Ce fichier est un panel, ce qui signifie que chaque déclarant est caractérisé par un identifiant invariant entre les années, ceci permettant de suivre les contribuables au cours du temps. Cet identifiant est le même que celui des fichiers POTE exhaustifs chainés. Ces fichiers, dont l'IPP dispose également, correspondent à l'exhaustif des déclarations à l'impôt sur le revenu (IR), en panel également, sur la période 2006-2017. Autrement dit, nous disposons pour chaque déclarant de ses informations relatives à l'ISF/IFI et à l'IR, ce qui nous donne une correspondance entre patrimoine taxable à l'ISF/IFI et revenus fiscaux.

2.1.3 Les données des fichiers de la taxe d'habitation et des propriétés bâties (FIDELI)

Le Fichier Démographique sur les Logements et les Individus (FIDELI) est une source de données produite par l'INSEE depuis 2015. Elle est constituée à partir de divers fichiers administratifs : le fichier de la taxe d'habitation, le fichier des propriétés bâties et le fichier des déclarations fiscales. Cette source de données est exhaustive : elle contient l'ensemble des logements et des individus en France. Dans TAXIPP 1.1, nous utilisons le millésime 2016. Ces données permettent de connaître de manière précise les ménages et leur localisation, ainsi que la valeur locative cadastrale de leur logement (utilisée dans le calcul de la taxe d'habitation).

2.1.4 L'enquête "Revenus Fiscaux et Sociaux" (ERFS)

L'enquête ERFS est produite par l'INSEE après appariement des données de l'Enquête Emploi en Continu (EEC), des fichiers de l'impôt sur le revenu et des données sur les prestations sociales reçues provenant de la CNAF, de la CNAV et de la CCMSSA. L'échantillon est constitué d'environ 50 000 ménages français. Dans TAXIPP 1.1, nous utilisons le millésime 2014. Nous utilisons l'information de l'ERFS relative aux liens entre les ménages, les foyers fiscaux et les familles, ainsi que les variables relatives au statut des individus sur le marché du travail (statut d'activité, temps de travail, secteur d'activité).

2.1.5 Les données de l'enquête "Budget de Famille" (BDF)

L'enquête « Budget de Famille » (BDF) est réalisée tous les 5 ans auprès d'environ 15 000 ménages et vise à étudier leur consommation. Elle renseigne pour chaque ménage les postes de dépenses (transport, logement, impôts, etc.) et les ressources (revenus, transferts sociaux, etc.). Le modèle TAXIPP 1.1 utilise le millésime 2011

et se restreint au champs de la France métropolitaine¹, pour un échantillon de 10 342 ménages. Ces données permettent d'estimer la composition des paniers de consommation des ménages et de simuler les taxes indirectes.

L'enquête BDF est utilisée afin de caractériser les effets redistributifs de la fiscalité indirecte par groupe de revenus. Il s'agit de l'approche employée dans Ben Jelloul et al. (2019a). Lorsque l'on souhaite évaluer les effets redistributifs horizontaux de la fiscalité indirecte (effets redistributifs au sein des groupes de revenus), l'enquête BDF est complétée par d'autres sources d'informations (Douenne, 2018b ; Douenne, 2018a)².

2.2 Fusion des données FELIN et ERFS

Le cœur du modèle TAXIPP 1.1 repose sur l'utilisation d'une base de données représentative de la population française obtenue suite à l'appariement statistique des données FELIN 2014 et ERFS 2014. La méthodologie utilisée pour apparier ces données est présentée en détails dans cette section. Les autres sources d'information présentées précédemment sont utilisées à part, pour simuler la taxe d'habitation et les taxes indirectes (voir le chapitre 3)³.

Un appariement statistique entre deux bases consiste à allouer à chaque individu d'une base de données (la base « receveuse ») un « plus proche voisin » au sein de

1. La fiscalité indirecte, et notamment celle portant sur les carburants, est fixée différemment hors de métropole.

2. L'hétérogénéité dans le niveau de consommation d'un bien donné au sein d'un groupe peut être primordiale. C'est le cas par exemple pour les dépenses d'énergie, qui dépendent, en plus du revenu, du type de logement, ou de sa localisation. De plus, les dépenses de l'enquête BDF sont répertoriées sur des durées courtes pour certains types de bien, ce qui surestime l'hétérogénéité au niveau individuel. Lorsque l'on s'intéresse par exemple aux effets redistributifs horizontaux de la fiscalité énergétique, nous procédons à un appariement statistique avec l'Enquête Nationale Transports et Déplacements (ENTD), pour corriger la distribution des dépenses au sein des groupes de revenu (voir Douenne 2018b pour la méthodologie employée). Lorsqu'on se focalise sur les effets redistributifs verticaux (entre groupes de revenus) de cette fiscalité (comme dans Ben Jelloul et al. (2019a)), un tel appariement n'est pas nécessaire. Si l'enquête BDF surestime la dispersion des consommations, elle permet d'obtenir les bonnes moyennes au sein de chaque groupe de revenus.

3. La version TAXIPP 2.0 (à venir) reposera sur l'utilisation d'une base de données unique, issue de l'appariement de l'ensemble de ces sources.

la deuxième base (la base « donneuse »). La notion de voisin le plus proche repose sur le calcul de la « distance » entre deux observations de chacune des deux bases de données. Cette distance est déterminée à partir des variables présentes dans les deux bases. Autrement dit, on alloue à chaque foyer fiscal de FELIN un foyer fiscal d'ERFS qui lui est proche, au regard des variables présentes dans les deux sources de données. Il s'agit donc de créer une base de données composée de foyers fiscaux fictifs, reproduisant les distributions jointes de l'ensemble des variables présentes dans les deux bases. L'objectif est de disposer d'une base unique comportant l'union de l'ensemble des variables des deux bases de données.

L'échantillon FELIN constitue la base de données « receveuse » dans notre processus d'appariement. Cette source étant plus précise que l'ERFS concernant les variables fiscales, elle constitue la source d'information centrale. Lorsqu'une variable est présente à la fois dans FELIN et dans l'ERFS, la priorité est donnée à l'information provenant de FELIN. Notre échantillonnage de référence est également l'échantillonnage de FELIN, celui d'ERFS étant appliqué seulement pour les variables d'ERFS non présentes dans FELIN.

Les données FELIN contiennent déjà pour chaque foyer fiscal le détail de la déclaration de revenus, c'est-à-dire l'ensemble des informations sur les revenus, la composition familiale du foyer fiscal, et l'ensemble des informations relatives aux crédits et réductions d'impôt. Les variables de l'ERFS qui nous intéressent, car non contenues dans FELIN sont :

- **la correspondance foyer fiscal - famille**, une famille n'étant pas forcément regroupée dans un seul foyer fiscal, comme c'est le cas pour les couples en concubinage par exemple ;
- la correspondance foyer fiscal - ménages, un ménage étant un ensemble d'individus vivant dans le même logement ;
- **les variables de l'enquête emploi sur le statut sur le marché du travail**,

utilisées pour la simulation des cotisations sociales (statut d'activité, temps de travail, secteur d'activité) ;

- **les revenus du capital** non présents dans la déclaration de revenus, et imputés dans l'ERFS ;
- **les loyers et les valeurs locatives des propriétaires occupants.**

2.2.1 Harmonisation préalable des deux bases

Dans un premier temps, les bases de données des foyers fiscaux FELIN et ERFS sont harmonisées et retraitées en vue de leur appariement. Nous produisons pour FELIN et pour l'ERFS une base avec une observation par foyer fiscal⁴, et un ensemble de variables ayant des définitions harmonisées entre les deux sources. La base de données de l'ERFS 2014 que nous utilisons pour l'appariement est la table foyer14, qui est la table contenant les informations de l'ensemble des déclarations fiscales des répondants de l'enquête emploi retrouvés.

Les revenus catégoriels sont reconstitués à partir des informations des déclarations fiscales, en vue de leur utilisation pour l'appariement. Les revenus sont divisés en quatre grande catégories, reprenant la décomposition utilisée dans l'ERFS⁵ :

- **les revenus salariaux au sens large.** Ils sont calculés comme la somme des revenus d'activité, des autres revenus imposables (chômage, préretraite), des salaires exonérés des agents d'assurance, et des indemnités de fonction des élus locaux.
- **les pensions au sens large.** Elles sont calculées comme la somme des pensions, des retraites, des rentes, des pensions alimentaires reçues, des pensions d'invalidité et des rentes viagères à titre onéreux.

4. Il existe des cas de double déclaration lorsque certains événements particuliers interviennent au cours de l'année. Les deuxièmes déclarations sont supprimées des deux bases. La double déclaration est obligatoire dans le cas du décès d'un des membres du couple et optionnelle dans le cas d'un enfant atteignant sa majorité.

5. Cf. Bilan de Production de l'ERFS 2014, p.12.

- **les revenus non salariaux.** Ils sont calculés comme la somme des revenus agricoles, des revenus industriels et commerciaux et des revenus non commerciaux.
- **les revenus du patrimoine.** Ils sont calculés comme la somme des revenus des valeurs et capitaux mobiliers (soumis ou non au prélèvement libératoire) et des revenus fonciers nets des déficits fonciers.

Les variables socio-démographiques (âge des membres du foyer, statut marital, nombre et type des personnes à charge) sont également harmonisées en vue de leur utilisation lors de l'appariement. Certaines valeurs d'âge sont manquantes dans les déclarations fiscales, ce qui nous oblige à faire des retraitements⁶.

2.2.2 Appariement statistique des foyers fiscaux

L'appariement statistique, ou « matching », vise à attribuer à chaque foyer fiscal FELIN un « plus proche voisin » parmi les foyers fiscaux de l'ERFS. La base « receveuse » est la base FELIN : cela signifie que c'est l'échantillonnage de FELIN qui sera reproduit *in fine*. La base « donneuse » est la base ERFS : elle est simplement utilisée pour compléter les informations non présentes dans FELIN. Lorsqu'une variable est présente dans chacune des deux bases, c'est la valeur dans FELIN qui sera utilisée lors des simulations. Le matching se fait sans remise (chaque foyer ERFS est choisi une seule fois comme plus proche voisin d'un foyer FELIN) afin de reproduire la distribution des variables récupérées dans l'ERFS.

6. La variable d'âge est parfois manquante dans les deux bases pour les personnes à charge majeures. Une personne à charge majeure peut être déclarée, sans que sa date de naissance figure dans la déclaration de revenus. Ceci s'explique principalement par la structure du formulaire n°2042 de la déclaration de revenus qui n'a prévu que deux champs pour indiquer nom, prénom, lieu et date de naissance des enfants majeurs. Dès lors, à partir du troisième enfant majeur, il est possible que les déclarants omettent tout simplement de renseigner cette information. En comparant la pyramide des âges issue des données FELIN avec la pyramide des âges de l'INSEE, il est possible de quantifier la proportion de valeurs manquantes âge par âge. À partir de cette information, on tire aléatoirement un âge à chaque individu concerné par ce problème en utilisant la distribution des âges manquants.

Appariement exact sur certains critères. Les foyers des deux sources de données sont appariés en fonction de leur ressemblance selon divers critères. Pour certains critères jugés primordiaux, l'appariement se fait de manière exacte, c'est-à-dire que deux foyers ne peuvent être appariés ensemble si ils ne sont pas identiques selon ces critères. Les dimensions sur lesquelles l'appariement est exact sont les suivantes :

- **Statut marital (en 2 catégories).** On distingue ici les foyers composés d'un couple (le statut marital déclaré est « marié » ou « pacsé ») et les foyers de célibataires (le statut marital déclaré est « célibataire », « veuf » ou « divorcé »).
- **Nombre d'enfants (en 5 catégories).** On distingue les foyers selon leur nombre de personnes à charge déclarées : aucune, un, deux, trois, quatre et plus.
- **Appartenance du foyer à la partie exhaustive de l'échantillon ou non (2 catégories).** Empiriquement, on constate que le seuil de revenu imposable à partir duquel l'échantillonnage devient exhaustif dans FELIN est de 170 000 euros environ. Les foyers ERFs et FELIN sont répartis en deux catégories selon que leur revenu imposable est inférieur ou non à ce seuil. La volonté d'apparier de manière exacte sur cette catégorie repose essentiellement sur le fait que cela allège les calculs notamment lors de l'opération de duplication des observations (cf. explications fournies ci-après)⁷.

Concrètement, apparier de manière exacte sur ces critères revient à effectuer l'appariement non pas sur la population entière des foyers ERFs et FELIN mais séparément au sein de chaque sous-catégorie de population. Ces sous-catégories sont définies par intersection des modalités des trois critères présentés. Il y a donc au total 18 sous-catégories.

7. Pour la catégorie des célibataires de la partie exhaustive, on ne divise l'échantillon qu'en trois sous-catégories de nombre de personnes à charge (aucune, 1, 2 et plus), du fait d'un nombre limité d'observations.

Recalage de la population ERFS. Le champ ERFS est plus restreint que celui de FELIN, car il exclut les ménages non ordinaires (individus vivant dans des structures communes telles que les maisons de retraites, les résidences étudiantes, les foyers de travailleurs, etc.) ainsi que les ménages vivant en France non métropolitaine. Suite aux retraitements évoqués précédemment, la base ERFS contient 32 377 066 foyers fiscaux, tandis que la base FELIN en contient 36 390 116. Le matching étant sans remise, les pondérations ERFS sont recalées pour obtenir le même nombre de foyers fiscaux que dans FELIN. L'appariement se faisant séparément pour chaque sous-catégorie définie précédemment, les pondérations ERFS sont recalées pour obtenir le même nombre de foyers entre les deux bases, dans chaque sous-catégorie.

Duplication des bases de données. Avant de procéder à l'appariement, il est nécessaire de dupliquer les foyers fiscaux afin de se ramener à une pondération commune pour toutes les observations. La duplication se fait au sein de chaque sous-catégorie de population. Pour les tables de la partie « exhaustive » de l'échantillon, les foyers sont dupliqués pour se ramener à une pondération commune de 1. Pour les tables de la partie « non exhaustive » de l'échantillon, les pondérations sont de 7,58⁸.

Appariement des foyers fiscaux. Au sein de chaque sous-catégorie, les foyers fiscaux FELIN sont appariés aux foyers fiscaux ERFS sur la base de la méthode d'appariement par score de propension. Le score de propension est ici calculé à partir des critères suivants :

- **Montant de l'impôt net.** L'utilisation de cette variable permet de capter l'hétérogénéité qui existe entre foyers, à revenu imposable donné, du fait notamment des réductions ou crédits d'impôt mobilisés.

8. Pondération minimale parmi les observations FELIN qui ne font pas partie de la partie exhaustive de l'échantillon

- **Montant des revenus catégoriels du déclarant (et du conjoint le cas échéant).**

Les quatre revenus catégoriels pris en compte sont ceux mentionnés précédemment.

- **Classe d'âge du déclarant (et du conjoint le cas échéant).** Les classes d'âge sont définies par tranche de cinq années.

La méthode du score de propension est généralement utilisée dans les évaluations d'impact pour constituer un groupe de contrôle comparable à un groupe d'individus traités. Le score de propension d'une observation s'interprète alors comme la probabilité d'être traitée conditionnellement à certaines caractéristiques observées. Ici, le score de propension est défini avec pour variable de « traitement » une dummy égale à 1 s'il s'agit d'une observation FELIN, et 0 s'il s'agit d'une observation ERFS. Le score de propension n'a donc pas d'interprétation particulière. Il sert à calculer un indice unique à partir de plusieurs variables pour évaluer la ressemblance de deux foyers fiscaux au vu de ces caractéristiques⁹.

2.2.3 Création de tables individus, foyers fiscaux, familles et ménages

Les dispositifs socio-fiscaux dans le système français se calculent au niveau de différentes « entités » : certains dispositifs sont calculés à l'échelle du ménage (comme les aides au logement), d'autres au niveau de la famille (comme les prestations sociales), d'autres au niveau du foyer fiscal (comme l'impôt sur le revenu) ou encore au niveau de l'individu (comme les cotisations sociales). Chaque variable au sein du calculateur OpenFisca a donc une entité de définition (cf. chapitre 3, section 3.1). Suite à l'appariement statistique FELIN-ERFS qui s'effectue au niveau des foyers fiscaux, il est nécessaire de reconstituer des tables relatives à chacune des

9. La version 2.0 de TAXIPP, en cours de développement, reposera sur l'utilisation de distances de Mahalanobis, plus précis, mais plus lourd (du fait du calcul d'une distance pour chaque couple d'observation FELIN et observation ERFS) et donc nécessitant davantage d'optimisation.

autres entités. Quatre tables sont produites afin d'être utilisées par le simulateur.

- **La table « individus ».** Cette base de données contient à la fois des variables FELIN et ERFS. Une fois les foyers fiscaux FELIN et ERFS appariés, il faut donc appairer les individus au sein de ces foyers entre eux. Le déclarant principal d'un foyer FELIN est apparié au déclarant principal du foyer ERFS apparié (sur la base de la variable de statut au sein du foyer). Il en va de même pour le conjoint, le cas échéant. Pour les personnes à charge, elles sont classées selon leur âge puis appariées entre elles en prenant en compte ce classement (ex : soit un foyer fiscal FELIN et un foyer fiscal ERFS matchés ensemble, et ayant chacun deux personnes à charge déclarées, on matche le plus jeune (respectivement le plus âgé) du foyer fiscal FELIN avec le plus jeune (respectivement le plus âgé) du foyer fiscal ERFS). Les informations individuelles relatives à l'âge, au statut marital et aux revenus individuels sont celles issues de la base FELIN. Les autres variables individuelles sont issues de l'ERFS.
- **La table « foyers fiscaux ».** Elle renseigne, pour chaque foyer fiscal, l'ensemble des cases de la déclaration de l'impôt sur le revenu, telles que renseignées dans FELIN.
- **La table « ménages ».** Elle comprend les informations relatives au logement (statut d'occupation du logement, loyer, loyer fictif). Il s'agit de variables issues de l'ERFS.
- **La table « familles ».** Par hypothèse, l'entité « famille » est supposée incluse dans l'entité « ménage ». Autrement dit, deux individus de la même famille appartiennent toujours au même ménage. La reconstitution des familles se fait en utilisant les variables de liens familiaux entre personnes d'un même ménage, selon la méthodologie de l'enquête ERFS¹⁰. Un individu peut être considéré comme personne à charge jusqu'à 24 ans.

10. Cf. Guide d'utilisation de l'Enquête Revenus Fiscaux 2002 rétropolée, p.135.

Un des intérêts de l'appariement FELIN-ERFS est d'allouer les foyers fiscaux à des ménages car les données FELIN ne renseignent pas sur cette entité. Dans un premier temps, nous allouons à un foyer fiscal ERFS (et donc au foyer fiscal FELIN avec lequel il a été matché) le ménage dans lequel se trouve le déclarant principal. Néanmoins, certains individus peuvent vivre dans un logement différent de celui du déclarant fiscal auquel ils sont rattachés. C'est le cas par exemple des individus de moins de 21 ans (ou 25 ans s'ils sont étudiants) ayant leur propre logement, mais rattachés fiscalement à leurs parents. Nous faisons donc des retraitements en aval de l'appariement, visant à prendre en compte l'appartenance d'un individu à un ménage différent de celui du déclarant principal auquel l'individu est rattaché fiscalement.

Il y a certains individus répondants à l'Enquête Emploi, et donc dans le champ de l'ERFS, dont on ne trouve pas de déclaration fiscale associée à leur adresse. La méthodologie de production de l'ERFS consiste à faire des imputations des revenus manquants pour ces individus, ou bien à les sortir de l'échantillon et à redresser en conséquence les pondérations (voir le bilan de production de l'ERFS 2014). Nous désignons ci-dessous par « table finale ERFS » cette table issue de ces imputations et redressements.

La base de l'ERFS utilisée pour l'appariement est la table *foyer14*, qui est, non pas la table finale ERFS, mais la table des déclarations fiscales des répondants à l'enquête emploi qui ont été retrouvées. Cette table *foyer14* est la table la plus comparable aux fichiers FELIN. Par conséquent, dans les ménages composés de foyers fiscaux où sont déclarés des individus ne vivant pas dans le ménage, les foyers fiscaux utilisés dans l'appariement contiennent des individus « en trop », par rapport au nombre d'individus dans ces ménages, tel que renseigné dans la table finale ERFS.

Pour chaque ménage de l'ERFS présent dans la table finale ERFS, on attribue à chaque individu de ce ménage un individu présent dans les déclarations fiscales

de la table `foyer14` auquel le ménage est associé. On fait cette attribution en fonction du statut de l'individu (déclarant, conjoint, personne à charge), et, pour les personnes à charge, en fonction de l'âge. On se retrouve, du fait des personnes à charge rattachées mais hors du ménage, avec un excès de personnes à charge pour certains ménages dans la table `foyer14`.

Ces personnes à charge « en trop », parce qu'ils sont présents dans la table `foyer14` de l'ERFS, ont chacun un individu FELIN avec lequel ils ont été matchés. L'exercice consiste à allouer à ces individus, non pas le ménage de leur déclarant fiscal, mais un autre ménage. Nous qualifions ci-dessous ces individus les « PAC hors ménage ».

Dans la table finale ERFS, on isole les étudiants de moins de 26 ans ayant leur propre ménage, et pour lesquels aucune déclaration fiscale n'a été retrouvée. Il s'agit des individus pour lesquels la variable `quelfic` de l'ERFS vaut « EE_NRT » (voir le bilan de production de l'ERFS 2014, p.101). On alloue les « PAC hors ménage » à ces individus de l'ERFS qui n'ont pas de déclaration. L'allocation se fait de manière aléatoire, conditionnellement à l'âge (on ne rapproche que deux individus ayant exactement le même âge). Puis, on récupère pour chacun des « PAC hors ménage » les variables du ménage de l'individu ERFS avec lequel il a été rapproché, seulement si ces variables ne sont pas déjà présentes dans FELIN. On suppose également que ces « PAC hors ménage » forment, non seulement un ménage à part, mais aussi une famille à part.

La duplication des bases de données décrite précédemment permet de se ramener à la plus petite pondération commune. Nous attribuons cette même pondération à l'ensemble des entités. Nous retombons donc dans une situation où la pondération d'un individu est la même que la pondération de son foyer fiscal, de sa famille et de son ménage. En bout de course, nous fusionnons les observations qui constituent des purs doublons, en additionnant leurs pondérations. Nous avons au final une table individus de 2 592 632 observations, une table foyers fiscaux à 1 387 934

observations, une table familles à 1 150 629 observations et une table ménages à 987 854 observations.

2.2.4 Vieillesse de la base FELIN-ERFS

La base de données produite à l'issue de l'appariement est relative à l'année 2014. Afin de pouvoir simuler le système socio-fiscal français jusqu'à l'année courante, cette base est vieillie jusqu'aux années postérieures¹¹. Ce vieillissement consiste à appliquer à la base les principales évolutions agrégées depuis 2014 en termes économiques et socio-démographiques.

TABLEAU 2.1 – Principaux agrégats utilisés pour le vieillissement

| | 2017 | 2018 | 2019 | 2020 |
|------------------------------------|-------------|-------------|-------------|-------------|
| Population totale | 66 987 103* | 67 270 809* | 67 548 018* | 67 818 998* |
| PIB nominal (en milliards d'euros) | 2 293,1 | 2 353,1 | 2 411,9* | 2 472,2* |
| PIB nominal par tête (en euros) | 34 232,0 | 34 979,5 | 35 706,8* | 36 453,3* |

NOTES : Les agrégats suivis d'un astérisque sont des prévisions. Au moment de la conception du modèle, la dernière année couverte était 2016 pour le recensement, et 2018 pour la comptabilité nationale.

SOURCES : Projections de population 2013-2070, INSEE. Mesures du PIB national, Comptabilité nationale (INSEE). Prévisions de taux de croissance du PIB nominal, OCDE (prévisions en date de septembre 2019).

Le nombre d'individus est recalé sur les évolutions de la population totale française à partir des données du recensement ou des projections de population de l'INSEE, par un ajustement uniforme des pondérations. Les revenus déclarés à l'impôt sur le revenu sont vieillies en utilisant les dénombrements fiscaux publiés par la DGFIP, qui renseignent pour chaque case de la déclaration de revenus le montant agrégé déclaré au niveau national. Pour les années les plus récentes pour lesquelles ces dénombrements ne sont pas disponibles, les revenus sont calés sur le taux de croissance du PIB nominal par tête observé ou projeté (voir le tableau 2.1)¹². Pour

11. Lors de la publication de ce guide méthodologique, nous vieillissons les bases de 2015 à 2020.

12. Seules les cases contenant au niveau agrégé les montants les plus importants sont vieillies. Il s'agit des cases 1AJ à 1DJ, 1AP à 1DP, 1AK à 1DK, 1AS à 1DS, 1AT, 1BT, 1AZ à 1DZ, 1AO à 1DO, 1AW à 1DW, 2DH, 2EE, 2DC, 2FU, 2CH, 2TS, 2TR, 2FA, 2CG, 2BH, 2CA, 4BE, 4BA, 4BB, 4BC, 4BD, 4BF, 6DE, 6GI, 6GJ, 6EL, 6EM, 6GP, 6GU, 6DD. Nous vieillissons également un ensemble de cases de la déclaration 2042-PRO, non pas sur les dénombrements mais sur la croissance du PIB nominal

les revenus non-déclarés mais imputés dans l'ERFS (revenus de l'assurance-vie, livrets d'épargne exonérés, etc.), le vieillissement se fait en utilisant le taux de croissance du PIB nominal par tête réalisé ou prédit (voir le tableau 2.1). Nous appliquons ce même vieillissement pour certaines prestations non simulées et dont nous injectons le montant tel qu'observé dans l'ERFS (voir le chapitre 3 pour la liste de ces prestations). Enfin, les loyers et loyers fictifs sont vieillis en utilisant les variations de l'indice de référence des loyers (IRL) publié par l'INSEE.

par tête, du fait de la plus grande volatilité dans le temps de la définition de ces cases. Il s'agit des cases 5TA, 5TB, 5TE, 5HO, 5HC, 5IC, 5HF, 5KO, 5KP, 5KC, 5KI, 5LC, 5HA, 5KA, 5IA, 5ND, 5NG, 5HQ, 5QB, 5QC, 5QI, 5RC, 5HY.

CHAPITRE 3

PROGRAMMES DE SIMULATION

3.1 Connection avec OpenFisca

À partir de la base de données décrite dans le chapitre précédent, TAXIPP 1.1 simule l'ensemble du système socio-fiscal à l'aide du calculateur open source OpenFisca. OpenFisca est un calculateur libre et collaboratif, que l'IPP co-développe avec différents acteurs (<https://fr.openfisca.org/>). Ce calculateur vise à simuler l'ensemble du système socio-fiscal, que ce soit dans le cadre de simulations avec des données, ou dans le cadre d'API web.

Une simulation sur base de données sous OpenFisca passe par la définition d'un *survey scenario*. Il s'agit d'un objet qui définit conjointement à la fois la base de données source et le système socio-fiscal à simuler. La première étape dans la définition d'un *survey scenario* est l'import de la base de données utilisée, et sa mise en forme afin que la structure des variables corresponde à celle requise pour fonctionner avec le simulateur. On parle de « connexion » au calculateur OpenFisca. Nous appelons cette étape le *data builder*.

Dans le code d'OpenFisca, les dispositifs simulés, tout comme les inputs permettant leur simulation (revenus, caractéristiques socio-démographiques) prennent la forme de variables. Chaque variable est définie à l'échelle d'une entité donnée et

d'une période donnée. Il existe quatre entités dans le simulateur :

- l'individu ;
- le foyer fiscal ;
- la famille ;
- le ménage.

À titre d'exemple, le montant d'impôt sur le revenu est une variable définie au niveau du foyer fiscal, à l'échelle annuelle. Les prestations familiales sont définies au niveau de la famille, à l'échelle mensuelle.

Le *data builder* consiste à adapter les variables de la base de données appariée FELIN-ERFS à la structure des variables du calculateur OpenFisca. La plupart du travail effectué dans cette étape représente donc de la mise en forme des données. Deux étapes du *data builder* sont néanmoins centrales : la gestion des périodes (sous-section 3.1.1) et la construction des revenus bruts (sous-section 3.1.2).

3.1.1 La gestion des périodes

Les revenus renseignés dans la base FELIN sont annuels car ils correspondent aux revenus des déclarations fiscales. Or, certains dispositifs socio-fiscaux dépendent de revenus infra-annuels. Le calcul du RSA, par exemple, repose sur les ressources des trois mois précédents la demande. Les variables de revenus définies à l'échelle mensuelle dans le calculateur sont obtenues en allouant les revenus annuels de manière uniforme entre chacun des mois de l'année.

D'autres dispositifs dépendent des revenus des années antérieures. La plupart des prestations sociales sous conditions de ressources dépendent ainsi des revenus de l'avant-dernière année. Étant donné que nous n'avons pas d'informations pluri-annuelles, nous obtenons les revenus des années antérieures en les déflatant par la croissance du PIB nominal par tête, ou par les prévisions de cette croissance ¹.

1. Les prévisions de croissance utilisées sont les mêmes que celles utilisées pour le vieillissement

3.1.2 La reconstitution des revenus bruts

Les revenus renseignés dans la base FELIN correspondent à des revenus imposables. Les règles de calcul associées au revenu imposable sont dépendantes de la législation fiscale. Lors des simulations, il est important de partir des montants de revenus bruts qui correspondent aux valeurs contractuelles de ces revenus, indépendantes des changements législatifs. C'est pourquoi nous convertissons les revenus imposables de FELIN 2014 en revenus bruts, à partir de la législation de 2014.

Les salaires bruts. Ils sont reconstruits à partir des salaires imposables en inversant le barème habituellement appliqué pour passer du brut à l'imposable. Ainsi, le salaire brut correspond au revenu imposable augmenté des cotisations sociales salariés et de la CSG. Le calcul de la CSG se fait en tenant compte de la catégorie du salarié (privé, contractuel, fonction publique d'État, hospitalière ou territoriale mais aussi cadre ou non cadre), de la nature du contrat de travail (temps plein, temps partiel) et du volume d'heures travaillées dans le cas d'un temps partiel.

Plusieurs types de retraitements sont nécessaires afin de procéder au calcul des cotisations sociales et de la CSG. Les informations relatives au secteur d'activité et aux heures travaillées sont tirées des variables de l'Enquête Emploi en Continue (EEC), présentes dans l'ERFS. Nous calons les effectifs des différentes catégories de salariés ainsi obtenus de sorte à atteindre des cibles attendues d'effectifs de salariés cadres et de fonctionnaires des différentes fonctions publiques² Par ailleurs, les cotisations sociales des fonctionnaires peuvent avoir comme base le traitement indiciaire (pension civile) ou les primes (RAFP). Seul le salaire imposable total est renseigné. Il est donc nécessaire de faire une hypothèse sur le taux de prime, afin

de la base appariée FELIN-ERFS (voir chapitre 2)

2. Les effectifs de la fonction publique sont tirés des publications de l'Insee (Sénigout et Duval, 2016) Les cibles d'effectifs des salariés du secteur privé sont tirées des publications de l'Agirc-Arrco (Direction technique de l'Agirc-Arrco, 2016). La réallocation entre salariés cadres et non-cadres du secteur privé se fait en fonction des rémunérations observées : sont alloués en priorité à la catégorie des cadres (déficitaires) les salariés non-cadres (excédentaires) ayant les rémunérations les plus élevées.

de remonter au traitement indiciaire brut. Celui-ci est imputé de manière homogène sur l'ensemble de la fonction publique à partir de l'information fournie par les rapport annuels sur l'état de la fonction publique, publiés chaque année par la Direction Générale de l'Administration et de la Fonction Publique (DGAFP).

Les revenus de remplacement bruts. Ils sont calculés à partir des revenus imposables (pensions de retraites, allocations chômage etc.) en y ajoutant les montants de CSG. Selon le niveau de ressources du contribuable, il existe différents taux de CSG applicables sur les revenus de remplacement (exonération complète, taux réduit, taux plein). À partir du revenu fiscal de référence et du nombre de parts renseignés dans la base FELIN, nous reconstituons le taux de CSG applicable selon la législation en vigueur.

Depuis 2019, l'éligibilité aux taux réduits³ de CSG a été assoupli. Le passage au taux plein n'intervient que lorsque le revenu fiscal de référence dépasse un certain seuil sur deux années consécutives. Afin de simuler le taux de CSG selon cette nouvelle règle, il est désormais nécessaire d'introduire des variations inter-annuelles dans le revenu fiscal de référence. Le revenu fiscal de référence entre deux années consécutives est donc « bruité » de manière à reproduire le nombre officiels de retraités éligibles aux taux réduits tel qu'annoncé dans le Projet de loi de finances 2019⁴.

3.2 Simulation du système socio-fiscal

La base de données appariée FELIN-ERFS permet de simuler :

- l'impôt sur le revenu ;

3. Par ailleurs, un nouveau taux réduit a été introduit *de facto* suite à l'annonce de la suppression de la hausse de CSG taux plein pour certains retraités.

4. Est appliqué un bruit blanc gaussien, dont la variance est fixée de sorte à atteindre cette cible de retraités éligibles aux taux réduits.

- les cotisations sociales ;
- les prélèvements sociaux ;
- les prestations sociales (aides au logement, prestations familiales, minima sociaux) ;

Le code permettant de simuler l'ensemble de ces dispositifs est open source⁵. Les sous-sections ci-dessous précisent certaines hypothèses spécifiques à l'exercice de simulation sur les données FELIN-ERFS.

L'ISF/IFI, la taxe d'habitation et la fiscalité indirecte sont simulés dans des modules à part. L'ISF et l'IFI sont simulés à partir du fichier panélisté des déclarations ISF/IFI (sous-section 3.2.7), la taxe d'habitation est simulée avec FIDELI 2016 (sous-section 3.2.8), et la fiscalité indirecte avec l'enquête BDF 2011 (sous-section 3.2.9). Ces simulations sont ensuite intégrées en aval au cœur du modèle (sous-section 3.2.10).

Le modèle TAXIPP 1.1 simule les prélèvements obligatoires et transferts monétaires destinés aux ménages. Les prélèvements obligatoires soumis aux entreprises, tout comme les cotisations sociales employeurs, ne sont pas pris en compte dans le modèle. L'IPP dispose actuellement d'un simulateur de la fiscalité entreprises, fonctionnant sur données administratives (Bozio et al., 2018), mais qui ne communique pas avec le modèle TAXIPP 1.1. L'hypothèse sous-jacente est la non-incidence de la fiscalité des entreprises sur les ménages.

3.2.1 L'impôt sur le revenu

Le modèle simule l'impôt sur le revenu grâce aux informations des données FELIN, qui contiennent pour un échantillon de déclarations fiscales l'ensemble des cases de la déclaration de revenus. Nous pouvons donc simuler avec précision les revenus catégoriels, leur application au barème, ainsi que les réductions et crédits

5. <https://github.com/openfisca/openfisca-france>

d'impôt. Seuls les réductions et crédits d'impôt créés après l'imposition des revenus de 2014 ne sont pas simulés, étant donné que nous ne disposons pas des cases associées à ces dispositifs dans les fichiers FELIN 2014. Concernant le prélèvement à la source en vigueur à partir de 2019, l'impôt payé à la source est égal, dans nos simulations, à l'impôt simulé sur les revenus 2018 (avec application du barème voté dans la loi de finances pour 2019) multiplié par le taux de croissance prévisionnel du PIB nominal par tête⁶.

La loi de finances pour 2019 prévoit un dispositif d'exonération d'impôt sur le revenu des rémunérations des heures supplémentaires. Les données fiscales ne contiennent pas d'information sur ces rémunérations. Nous imputons à chaque individu une rémunération au titre des heures supplémentaires à partir de statistiques descriptives par quartile de salaire brut produits par la DARES⁷.

3.2.2 Les cotisations sociales

Les cotisations sociales applicables aux salariés sont simulées à partir des rémunérations brutes reconstituées (salaire brut dans le secteur privé, traitement indiciaire et primes dans le secteur public) et des informations utilisées pour cette reconstitution (cf. sous-section 3.1.2). Pour les salariés relevant du secteur privé, sont calculées, durant les périodes où elles s'appliquent, les cotisations AGFF, Agirc-Arrco, APEC, chômage, maladie (incluant maternité, invalidité et décès), vieillesse (plafonnée et déplafonnée) et la cotisation exceptionnelle temporaire. Pour les salariés relevant du secteur public, sont calculées également la cotisations pour la pension civile, celle du régime additionnel de la fonction publique (RAFP), la cotisation du régime complémentaire des contractuels (IRCANTEC) et la contribution

6. Les prévisions de croissance utilisées sont les mêmes que celles utilisées pour le vieillissement de la base appariée FELIN-ERFS (voir chapitre 2)

7. Source : « Les salariés effectuant des heures supplémentaires rémunérées en 2010 », DARES Analyses n°15, mars 2013, tableaux 2 et 3. Nous utilisons une probabilité de recours aux heures supplémentaires par quartile de salaire de base, ainsi qu'une rémunération moyenne au titre de ces heures par quartile de salaire de base, parmi les recourants à ces heures. Ces chiffres ont été produits avec l'enquête sur le coût de la main d'œuvre et la structure des salaires (ECMOSS) de 2010.

exceptionnelle de solidarité. Pour la simulation des exonérations de cotisations sociales sur les heures supplémentaires entrant en vigueur en 2019, nous utilisons les mêmes imputations de rémunération des heures supplémentaires que pour l'impôt sur le revenu.

3.2.3 Les prélèvements sociaux

Le modèle simule la CSG et la CRDS sur les revenus d'activité et de remplacement à partir des revenus bruts estimés dans le *data builder* (cf. sous-section 3.1.2). Nous simulons également l'ensemble des prélèvements sociaux sur les revenus du patrimoine et les revenus de placement. Pour ces revenus, la base fiscale des prélèvements sociaux est plus large que la base fiscale soumise à l'impôt sur le revenu. Nous ajoutons aux revenus du capital renseignés dans les déclarations fiscales les revenus d'assurance-vie et les intérêts des plans et comptes épargne logement, tels qu'imputés dans l'ERFS.

3.2.4 Les prestations familiales

Nous simulons les prestations familiales suivantes : les allocations familiales (AF), le complément familial (CF), l'allocation de rentrée scolaire (ARS), l'allocation de base de la Prestation d'accueil du jeune enfant (PAJE) et l'allocation de soutien familial (ASF).

Ces allocations sous conditionnelles aux ressources de l'avant-dernière année. Nous approximons ces ressources par le montant de ressources de l'année courante, déflatées par la croissance du PIB nominal par tête (voir la sous-section 3.1.1).

L'ERFS contient des informations sur les prestations effectivement perçues. Ainsi, nous injectons directement dans les données les prestations que nous ne simulons pas. Il s'agit de la Prestation partagée d'éducation de l'enfant (PreParE), destinée aux parents de jeunes enfants réduisant ou cessant leur activité professionnelle ; du

Complément de libre choix de mode de garde (CMG), qui est une prise en charge partielle des dépenses d'assistante maternelle ; de l'allocation d'éducation de l'enfant handicapé (AEEH). La simulation de la PreParE et du CMG nécessiterait d'avoir des informations sur les variations d'offre de travail suite à la naissance d'un enfant, et sur les choix de mode de garde. La simulation de l'AEEH nécessiterait des informations plus précises sur le handicap. Il est donc décidé de ne pas simuler ces dispositifs. En revanche, l'injection du montant de ces prestations renseigné dans l'ERFS permet de prendre en compte ces dispositifs dans le revenu disponible des ménages et dans les bases ressources des autres prestations lorsque cela est nécessaire.

3.2.5 Les aides au logement

Le modèle simule l'Allocation de logement familiale (ALF), l'Allocation de logement social (ALS) et l'Aide personnalisée au logement (APL). Tout comme les prestations familiales, les ressources de l'avant dernière année sont approximées par les ressources de l'année courante déflatées par l'évolution du PIB nominal par tête (voir la sous-section 3.1.1)⁸.

Les aides au logement sont destinées aux locataires et sont calculées en fonction du loyer. Les informations sur le statut d'occupation du logement et sur le loyer sont prises de l'ERFS. Nous faisons également l'hypothèse que tous les logements sont situés dans la zone APL 2.

8. Nous ne simulons pas pour le moment le passage à la prise en compte des ressources de l'année courante dans le calcul des aides au logement, qui interviendra à partir d'avril 2019. Néanmoins, dans le cadre des travaux sur l'évaluation du budget 2019 (Ben Jelloul et al., 2019a), nous imputons des pertes associées à cette réforme. Une solution aurait été d'estimer ces pertes en simulant les aides au logement avec la prise en compte des revenus de l'année courante, sans application du déflatage. Néanmoins, cette méthode n'aurait pas capturé les variations inter-annuelles de ressources qui ici sont cruciales pour analyser cette réforme. Pour imputer ces pertes, nous avons alloué de manière uniforme une perte totale de 0,9 milliards d'euros (correspondant aux prévisions gouvernementales, mentionnées dans le projet de loi de finances pour 2019, p.17) à l'ensemble des foyers ne percevant pas d'aides au logement d'après la simulation du système avant réforme. L'intuition derrière cette imputation est que les perdants à cette réforme sont potentiellement des ménages ayant aujourd'hui des ressources trop importantes pour percevoir des aides au logement, alors que leurs ressources de l'avant dernière année leur donne droit à ces aides.

3.2.6 Les minima sociaux

Le modèle simule le Revenu de solidarité active (RSA), la Prime d'activité (PA) et l'Allocation de solidarité aux personnes âgées (ASPA).

Les minima sociaux associés au handicap, à savoir l'Allocation adulte handicapé (AAH) et son complément (CAAH) ne sont pas simulés, pour les mêmes raisons que l'AEEH. En revanche, on injecte dans les données d'input les montants de ces dispositifs renseignés dans l'ERFS afin de prendre en compte ces allocations dans le revenu disponible des ménages. L'Allocation de solidarité spécifique (ASS) n'est également pas simulée. Cette aide, destinée aux personnes sans emploi ayant épuisé leurs droits aux allocations chômage, dépend de l'historique des individus sur le marché du travail⁹. Faute d'information sur cet historique, ce dispositif n'est pas simulé, ce qui tire potentiellement à la hausse le montant des dispositifs dont les bases ressources incluent cette aide.

Le RSA et la prime d'activité dépendent des ressources des trois mois précédents la demande de la prestation. Nous disposons de données de revenus annuelles et répartissons ces ressources de manière uniforme entre les mois (cf. sous-section 3.1.1). Cette règle de répartition infra-annuelle tend à sous-estimer les droits au RSA et à la prime d'activité pour des individus connaissant des variations importantes de revenus d'activité à l'échelle d'une année. En parallèle, l'existence de non-recourants à ces prestations tend à surestimer les dépenses au titre du RSA et de la PA, si l'on fait l'hypothèse dans nos simulations que l'ensemble des éligibles à ces prestations demandent à les percevoir.

Lorsque nous simulons le RSA et la prime d'activité sous l'hypothèse d'absence de non-recours, nous simulons des dépenses de RSA très proches de celles constatées dans les agrégats officiels. En 2016, les dépenses agrégées de RSA s'élevaient

9. Pour pouvoir bénéficier de l'ASS, il faut notamment avoir épuisé ses droits à l'Allocation d'aide au retour à l'emploi (ARE), et avoir travaillé au moins 5 ans durant les 10 ans précédant la fin du dernier contrat de travail.

à 11,2 milliards d'euros¹⁰. Sans imputation du non-recours, le modèle simule 9,7 milliards d'euros de dépenses. Nous supposons donc un recours de 100 % parmi les bénéficiaires du RSA. Pour les bénéficiaires de la prime d'activité, mais non-bénéficiaires du RSA, nous supposons un taux de non-recours de 27 %¹¹. Les non-recourants parmi cette population sont tirés aléatoirement. La probabilité de tirage est constante pour l'ensemble des éligibles à la prime d'activité et non-éligibles au RSA.

Le modèle simule également du non-recours à l'ASPA dont le calcul repose sur les revenus trimestriels, tout comme le RSA. Néanmoins, le public éligible à l'ASPA, âgé de plus de 65 ans, est beaucoup moins sujet à des variations de revenus dues à des entrées et sorties fréquentes dans l'emploi. De plus, si les ressources trimestriels d'un individu dépassent le seuil d'éligibilité, sont prises en compte les ressources des douze derniers mois. Simuler l'ASPA à partir de revenus annuels est donc moins sujet à des risques de sous-estimation des droits. En revanche, l'ASPA est également sujette à non-recours. Lorsque nous simulons ce dispositif sans imputer de non-recours, nous sur-estimons les masses d'ASPA. En 2016, la masse versée au titre du minimum vieillesse (anciens dispositifs et ASPA) était de 3,1 milliards d'euros¹² tandis que le modèle simule une dépense totale de 4,3 milliards d'euros. Nous choisissons donc de tirer aléatoirement des bénéficiaires parmi les familles éligibles afin de reproduire les dépenses officielles. La probabilité d'être choisi comme bénéficiaire au sein des éligibles est uniforme et égale à 25 %.

3.2.7 L'imposition du patrimoine

La nouveauté de la version 1.1 de TAXIPP est l'utilisation du fichier panélisté des déclarations ISF/IFI, apparié aux fichiers administratifs de l'impôt sur le revenu.

10. Source : CNAF, Dépenses tous régimes de prestations familiales et sociales 2016.

11. Ce chiffre est en ligne avec celui avancé lors de l'évaluation de la prime d'activité (Ministère des solidarités et de la santé (2017)).

12. Source : Fonds de Solidarité Vieillesse.

Dans la version 1.0, l'ISF et l'IFI étaient simulés à partir des revenus fiscaux. Le patrimoine total était estimé par la méthode de capitalisation : nous déduisons le patrimoine total des foyers fiscaux en appliquant aux revenus fiscaux l'inverse de taux de rendement du capital, tels qu'obtenus à partir de la comptabilité nationale. Le patrimoine taxable était ensuite obtenu à partir de méthodes de calage, afin d'obtenir une base fiscale cohérente avec les masses de patrimoine taxable telles que documentées dans les publications administratives (voir Ben Jelloul et al., 2019b).

Dans la version 1.1 du modèle TAXIPP, nous simulons l'ISF et l'IFI dans un module à part, à partir du fichier panélisté des déclarations ISF/IFI. Nous simulons l'impôt analysé pour chaque contribuable à partir de l'ensemble des informations de ces déclarations. Puis, nous apparions le fichier ainsi obtenu avec les fichiers POTE. Nous obtenons ainsi une distribution d'impôt, non pas pour la seule population des déclarants ISF/IFI, mais pour toute la population des contribuables à l'impôt sur le revenu. Cet appariement permet également de disposer d'une base de données comprenant simultanément l'information relative au patrimoine taxable et aux revenus fiscaux des foyers fiscaux. L'intégration de cette distribution, avec le cœur du modèle est décrite dans la sous-section 3.2.10.

L'utilisation de ces nouvelles sources administratives permet de faire des progrès substantiels dans la simulation de l'imposition sur le patrimoine. La méthodologie employée dans TAXIPP 1.0 rencontrait deux principales limites. D'une part, la méthode de capitalisation ne prend pas en compte de manière fine l'hétérogénéité des taux de rendement des revenus du capital entre foyers fiscaux. D'autre part, cette méthodologie supposait un ratio entre patrimoine taxable et patrimoine total identique pour l'ensemble des redevables de l'ISF/IFI. Autrement dit, nous supposions que les foyers fiscaux étaient impactés de manière "uniforme" par les "niches" de l'ISF/IFI¹³. Les fichiers administratifs de l'ISF/IFI permettent de s'affranchir de ces

13. Par niche de l'ISF, nous désignons l'ensemble des réductions d'impôt, ainsi que l'ensemble des éléments du patrimoine des ménages hors du champ du patrimoine taxable (comme par exemple les biens professionnels pour l'ISF).

hypothèses, en renseignant la distribution jointe des revenus fiscaux et du patrimoine taxable. L'utilisation de ces fichiers ne permet pas en revanche de simuler des impôts ayant pour patrimoine taxable un spectre plus large du patrimoine total des ménages.

3.2.8 La taxe d'habitation

La taxe d'habitation est simulée avec la base FIDELI 2016. Cette base contient, pour l'exhaustif des logements, la valeur locative cadastrale, et des informations sur les revenus fiscaux et la composition familiale au sein de chaque logement. Cette simulation se fait dans un module séparé (voir la sous-section 3.2.10).

Le modèle simule la taxe d'habitation perçue par la commune, et celle perçue par les établissements publics de coopération intercommunale (EPCI). Nous prenons en compte les dispositifs d'exonération, de plafonnement, et d'abattements, sauf l'abattement spécifique au titre du handicap faute d'informations sur ce point. Les barèmes appliqués sont ceux issus des délibérations (taux de taxe, abattements) de 2016. Seule la taxe d'habitation sur les résidences principales est simulée. En bout de course, nous calons les masses de taxe d'habitation ainsi obtenues sur les produits agrégés officiels de taxe d'habitation tels que renseignés dans les rapports de l'Observatoire des finances et de la gestion publique locales (OFGL)¹⁴. Ces agrégats ne sont disponibles que jusqu'en 2017 à cette date. Pour les années suivantes, nous appliquons aux montants de taxe d'habitation des ménages le taux d'inflation prévisionnel, qui est également le taux de revalorisation des valeurs locatives cadastrales.

14. Pour 2017, la masse cible de taxe d'habitation est de 22,282 milliards d'euros : voir le rapport de 2018 de l'OFGL, p.185

3.2.9 La fiscalité indirecte

Les effets de la fiscalité indirecte sur le budget des ménages sont simulés à partir de l'enquête Budget de Famille (BDF) 2011. Les agrégats de la Comptabilité Nationale sont utilisés pour vieillir les données à l'année pour laquelle les simulations sont effectuées.

L'enquête BDF rapporte les consommations par type de bien, en suivant la nomenclature internationale COICOP. Pour chaque type de bien consommé, le modèle calcule ensuite les dépenses en fiscalité indirecte pour chaque ménage en appliquant les taux en vigueur. Cela permet de calculer les montants payés par les ménages en TVA, en TICPE, ainsi qu'en taxes sur le tabac et sur l'alcool. Le modèle permet également de simuler des réformes sur l'ensemble de ces taxes, ainsi que de nouveaux dispositifs hypothétiques. Une hypothèse importante lorsque l'on simule ces réformes de la fiscalité porte sur leur incidence sur les prix. Le modèle est flexible de ce point de vue. Il permet de considérer des incidences au cas par cas, en fonction de la nature des réformes étudiées, en empruntant des hypothèses d'incidence de la littérature.

Dans le cas des énergies, l'enquête renseigne les dépenses pour les carburants de transports mais ne distingue pas les dépenses en essence des dépenses en diesel. Ces deux combustibles étant soumis à des taux de taxation différents, il est nécessaire de les séparer *ex post*. L'enquête renseigne en revanche pour chaque ménage son nombre de véhicules diesel et son nombre de véhicules essence. De plus, les « Comptes des Transports » renseignent la dépense moyenne par véhicule essence et la dépense moyenne par véhicule diesel à l'échelle nationale¹⁵. Nous utilisons la différence entre ces deux dépenses moyennes pour allouer la dépense totale en carburant de transport d'un ménage pour chacun de ses véhicules tels que renseignés dans l'enquête. Nous obtenons donc pour chaque ménage ses dépenses en diesel et

15. Source : « Les Comptes des Transports en 2017 », 55^{ème} rapport de la Commission des comptes des transports de la Nation.

ses dépenses en essence, auxquelles nous pouvons appliquer la fiscalité respective de ces deux carburants.

Dans le cas d'accises, il est nécessaire de traduire ces dépenses en quantités pour appliquer les taux de fiscalité. Cette conversion, évidente pour la plupart des biens concernés, est plus complexe pour l'électricité et le gaz puisque ces deux énergies sont soumises à des contrats comprenant un prix fixe et un prix unitaire. Il est donc nécessaire d'imputer à chaque ménage un contrat spécifique pour déterminer la part fixe de ses dépenses, le prix unitaire correspondant, et déduire ainsi les quantités consommées. Pour cette imputation, nous supposons que les ménages n'utilisent que des contrats règlementés. Pour le gaz, nous attribuons à chaque ménage le type de contrat maximisant sa consommation compte tenue de sa dépense renseignée dans l'enquête (les ménages ayant des dépenses élevées se verront imputer un contrat avec une part fixe plus importante et un prix marginal plus faible). Pour l'électricité, le type de contrat dépend de la capacité du compteur du ménage. Nous utilisons des informations externes sur le nombre de compteurs au niveau national par type de compteur. Puis, nous reproduisons cette distribution de compteurs en imputant les compteurs ayant les plus grosses capacités aux ménages renseignant les dépenses les plus importantes. Puis, nous déduisons pour chaque ménage, à partir de son type de compteur, son type de contrat.

Afin de ne pas surestimer l'incidence des réformes fiscales sur le budget des ménages, le modèle prend en compte leurs réactions aux changements de prix. Ces comportements sont intégrés via des élasticités. Pour les énergies, elles sont obtenues via un modèle de demande estimé sur l'enquête BDF (voir Douenne 2018b pour la méthodologie employée). Les élasticités sont hétérogènes en fonction du niveau de vie des ménages et de leur localisation géographique. En moyenne, l'élasticité prix non compensée est de -0,45 pour les carburants de transports, et de -0,2 pour l'énergie du logement. Pour le tabac, nous supposons une élasticité de -0,5, d'après les résultats de Hill et Legoupil (2018).

3.2.10 Assemblage des différents modules

La base appariée FELIN-ERFS, connectée au calculateur OpenFisca permet de simuler l'ensemble des dispositifs socio-fiscaux mentionnés précédemment à l'exception de la taxe d'habitation et de la fiscalité indirecte, simulées à part sur des données spécifiques. La taxe d'habitation est simulée avec les données FIDELI 2016, tandis que la fiscalité indirecte est simulée avec l'enquête BDF 2011.

L'ensemble de ces simulations nous permettent notamment de produire des statistiques sur les effets moyens de l'ensemble du système socio-fiscal par centile de niveau de vie¹⁶. Pour cela, nous produisons tout d'abord ces effets moyens avec le cœur du modèle, c'est-à-dire les simulations faites via la base appariée FELIN-ERFS. Puis, nous produisons avec la base FIDELI 2016 des effets moyens relatifs à la taxe d'habitation par centile de revenus fiscaux catégoriels par unité de consommation¹⁷. FIDELI ne contient pas directement tous les éléments du système redistributif, raison pour laquelle nous ne classons pas les ménages par niveau de vie. Puis, nous superposons les effets moyens par centile hors taxe d'habitation (obtenus avec le cœur du modèle) avec les effets moyens par centile relatifs à la seule taxe d'habitation (obtenus avec FIDELI 2016). L'hypothèse sous-jacente est que les ménages d'un centile donné de niveau de vie se trouvent dans le même centile de revenus fiscaux catégoriels par unité de consommation (ex : les ménages situés dans le 1^{er} centile de niveau de vie sont dans le 1^{er} centile de revenus fiscaux catégoriels par unité de consommation). Autrement dit, nous supposons que le passage des revenus fiscaux catégoriels par unité de consommation au niveau de vie ne change pas

16. Voir Ben Jelloul et al. (2019a) pour un exemple. Un centile est un seuil à partir duquel un individu se situe dans une catégorie de population. Par exemple, le 1^{er} centile de niveau de vie est le seuil de niveau de vie en-dessous duquel un ménage est situé parmi les 1 % des ménages à plus faibles niveaux de vie. Ici, par abus de langage, nous appelons « centiles » des catégories de ménages : le 1^{er} « centile » de niveau de vie est le groupe des 1 % des ménages ayant les niveaux de vie les plus faibles, le dernier correspondant aux 1 % des ménages ayant les niveaux de vie les plus élevés.

17. Cette variable servant à définir les centiles correspond à la somme des revenus fiscaux catégoriels tels que renseignés dans FIDELI 2016, que l'on divise par le nombre d'unités de consommation.

le classement des ménages.

Nous faisons une opération similaire de superposition avec la simulation de l'imposition sur le patrimoine, et avec la simulation de la fiscalité indirecte. Pour l'imposition sur le patrimoine, nous produisons des effets moyens par centile de revenu fiscal de référence par part fiscale. Pour la définition de ces centiles, l'unité est le foyer fiscal, et non pas le ménage. Pour la fiscalité indirecte, nous calculons des effets moyens par décile de niveau de vie, et non pas par centile, du fait de la taille limitée de l'échantillon de l'enquête BDF. Nous supposons donc que l'effet moyen est le même au sein de chaque centile d'un décile de niveau de vie. Puis, nous superposons ces effets moyens avec les autres effets moyens évoqués ci-dessus.

CHAPITRE 4

EVALUATIONS DE RÉFORMES

SOCIO-FISCALES

Le modèle TAXIPP est utilisé pour produire des évaluations *ex ante* de réformes socio-fiscales. Il permet notamment de simuler les effets budgétaires et redistributifs de réformes des prélèvements obligatoires et transferts monétaires. La section 4.1 de ce chapitre décrit la méthode de définition d'une réforme socio-fiscale dans le simulateur. La section 4.2 détaille les différents types de résultats produits par TAXIPP 1.1.

4.1 La définition d'une réforme

L'analyse des effets d'une réforme consiste à comparer un système « avant réforme » avec un « système réformé » (ou « réforme »). On distingue deux modalités de définition de ces deux systèmes.

- 1 On part d'un système en vigueur durant l'année N , qui constitue le système « avant réforme ». Puis, on apporte des modifications à ce système pour obtenir le système après réforme. Cette approche est par exemple utilisée lorsque l'on veut étudier une réforme structurelle du système socio-fiscal, à partir d'un

système en vigueur durant une année donnée. Il s'agit de l'approche adoptée dans Ben Jelloul et al. (2018), où sont définis différents scénarios de mise en place d'un revenu de base à partir de la législation de 2016.

2 Lorsque l'on souhaite évaluer *ex ante* des mesures socio-fiscales entrant en vigueur pour une année à venir, on définit le système émanant de ces mesures. Ce système constitue le système après réforme. Dans un second temps, on définit un « système contrefactuel », qui correspond au système avant réforme. Il s'agit du système qui serait en vigueur si aucune des mesures évaluées n'était adoptée¹. Cette approche est celle adoptée dans Ben Jelloul et al. (2019a). Les paragraphes ci-dessous décrivent les modalités de définitions d'un système contrefactuel.

La version 1.1 de TAXIPP contient un outil permettant de créer un système socio-fiscal contrefactuel fictif. Chaque année, les paramètres des dispositifs socio-fiscaux sont soumis à des revalorisations « par défaut », afin notamment de prendre en compte l'inflation. Un système contrefactuel défini pour l'année $N + 1$, à partir d'une année N , vise à appliquer au système en vigueur à l'année N ces revalorisations par défaut. Ainsi, la comparaison du système réformé avec ce système contrefactuel permet d'identifier les mesures socio-fiscales épurées de l'ensemble de ces revalorisations par défaut.

Le système socio-fiscal « en vigueur durant l'année N » correspond au système socio-fiscal relatif à l'ensemble des dispositifs dont le *paiement* par les ménages se fait durant cette année N . Il s'agit donc des prestations sociales perçues durant l'année N , et des prélèvements obligatoires payés durant cette année-là. L'impôt sur le revenu du système en vigueur durant l'année N est donc, jusqu'en 2018 (avant le prélèvement à la source), l'impôt payé en N au titre des revenus de l'année $N - 1$.

1. Soit N l'année courante. Soit un ensemble de mesures entrant en vigueur pour l'année $N + 1$. Le système contrefactuel est dans ce cas le système qui serait en vigueur en $N + 1$ si aucune nouvelle mesure n'entrerait en vigueur.

De cette définition, on définit un système contrefactuel pour l'année $N + 1$ à partir du système de l'année N , en procédant à deux étapes :

- **Création du système en vigueur au 31/12/N.** Prendre pour référence le 31/12/N implique que toute mesure entrant en vigueur en cours d'année N est hors du champ de l'évaluation en $N + 1$, même si ces mesures sont entrées en vigueur à la fin de l'année N .
- **"Vieillessement" du système jusqu'au 31/12/N+1.** Les paramètres monétaires qui font l'objet d'une revalorisation annuelle sont revalorisés à partir de l'inflation prévisionnelle pour l'année $N + 1$. Nous revalorisons les paramètres de l'impôt sur le revenu sur l'inflation de l'année N , du fait du décalage d'une année entre le paiement de cet impôt, et les revenus faisant l'objet de l'imposition². La revalorisation concerne, entre autres, les montants de base des prestations sociales, les seuils de revenus utilisés dans certains calculs de dispositifs socio-fiscaux, etc. Pour les dispositifs définis à l'échelle infra-annuelle, les revalorisations légales ont parfois lieu en cours d'année. La simulation infra-annuelle de ces dispositifs (voir la partie 3.1.1) permet de coder ces revalorisations en cours d'année, autrement dit de ne pas les appliquer à l'ensemble de l'année civile.

4.2 Les résultats produits

Une fois le système avant réforme et après réforme définis, il est possible de comparer les effets budgétaires et redistributifs de la réforme étudiée. Plusieurs

2. Cette correction est également appliquée pour le système contrefactuel de 2019, afin d'avoir une situation contrefactuelle où le prélèvement à la source ne serait pas entré en vigueur. Concernant la modélisation du prélèvement à la source dans le système en vigueur en 2019, nous prenons actuellement l'impôt simulé sur les revenus de 2018, que nous vieillissons avec un taux de croissance prévisionnel du PIB nominal par tête (voir le tableau 2.1). Cette approche est à améliorer, mais correspond néanmoins à la construction générale du prélèvement à la source qui consiste à appliquer aux revenus courants, pour le calcul des acomptes mensuels, un taux moyen de taxation résultant de la dernière déclaration de revenus.

types de résultats peuvent être produits.

Les cas-types. Il s'agit de définir un « type » de ménage, en fixant l'ensemble des caractéristiques utilisées pour le calcul des prélèvements obligatoires et des prestations sociales. Une approche souvent utilisée est de fixer l'ensemble de ces caractéristiques, et d'en faire varier une seule³. L'exercice consiste à comparer le résultat obtenu avec le système avant et le système après réforme. Ce type d'output est souvent utilisé lorsque l'on veut illustrer l'impact d'une réforme sur une catégorie particulière de la population. Il prend souvent la forme d'un graphique. Aucune base de données n'a besoin d'être utilisée pour la production de ce type de résultat. Voir p.65 de Ben Jelloul et al. (2018) pour un exemple.

Le coût budgétaire d'une réforme. Le coût budgétaire d'une réforme correspond à la différence entre la dépense publique nette des recettes publiques, simulée dans le système socio-fiscal réformé, et celle simulée dans le système socio-fiscal avant réforme. Un coût total positif (respectivement négatif) signifie que le système réformé engendre une hausse (respectivement une baisse) de la dépense publique (nette des recettes).

Les effets redistributifs. Le modèle TAXIPP permet de caractériser les gains et les pertes des ménages suite à une réforme socio-fiscale, c'est-à-dire à identifier les gagnants et les perdants en fonction d'une caractéristique donnée, ainsi que les montants des gains et des pertes (ex : montants de gains et de pertes moyennes par catégorie de revenus ; voir Ben Jelloul et al. (2019a), graphique 1 pour un exemple).

3. Par exemple : soit un ménage composé d'un célibataire sans enfant. Comment ses prélèvements obligatoires et ses prestations sociales varient-elles en fonction de ses revenus du travail ?

CONCLUSION GÉNÉRALE

Ce guide méthodologique décrit le fonctionnement de l'outil de microsimulation TAXIPP dans sa version 1.1. Développé par l'Institut des politiques publiques (IPP), cet outil vise à simuler à partir de diverses sources administratives le système socio-fiscal français. Il permet d'évaluer les impacts budgétaires et redistributifs de dispositifs socio-fiscaux actuels ou de réformes. L'apport de la version 1.1, par rapport aux versions précédentes, est l'utilisation des fichiers administratifs de l'ISF et de l'IFI, qui permettent d'améliorer de manière importante la simulation de l'imposition sur le patrimoine. TAXIPP 1.1 fait suite à la sortie récente de la versions 1.0, caractérisée par plusieurs changements majeurs. L'utilisation d'une base de données inédite issue de l'appariement statistique des fichiers de l'impôt sur le revenu (échantillon FELIN) et des données de l'enquête Revenus Fiscaux et Sociaux (ERFS) améliore la finesse de l'échantillon et la fiabilité des simulations. L'utilisation des données FIDELI, représentant l'exhaustivité des logements, avec leur géolocalisation, permet de simuler la taxe d'habitation et d'intégrer la dimensions territoriale dans l'analyse de cet impôt. Par ailleurs, depuis la version 1.0, TAXIPP simule la législation socio-fiscale à partir du calculateur libre et collaboratif, OpenFisca.

La version 2.0 du modèle est en cours de développement. La base de données utilisée pour les simulations sera issue d'un appariement statistique entre les données FIDELI, les fichiers FELIN et les fichiers Postes des déclarations annuelles des données sociales (DADS Postes). Cette base appariée aura pour échantillonnage principal celui de FIDELI, à savoir un exhaustif des ménages français. L'objectif est

de pouvoir analyser de manière précise des effets redistributifs entre territoires. L'utilisation des DADS, venant remplacer les données de l'Enquête Emploi améliora la qualité des simulations des cotisations salariales, pour lesquelles de nombreuses imputations décrites dans ce guide ne seront plus nécessaires.

RÉFÉRENCES

- Bach, L., Bozio, A., Fabre, B., Guillouzouic, A., Leroy, C., et Malgouyres, C. (2019a). Evaluation d'impact de la fiscalité des dividendes. *Rapport de l'IPP*, (25).
- Bach, L., Bozio, A., Fabre, B., Guillouzouic, A., Leroy, C., et Malgouyres, C. (2019b). Quelles leçons tirer des réformes de la fiscalité des revenus du capital? *Note de l'IPP*, (46).
- Ben Jelloul, M., Bozio, A., Cottet, S., Fabre, B., et Leroy, C. (2018). Revenu de base : simulations en vue d'une expérimentation. *Rapport de l'IPP*, (18).
- Ben Jelloul, M., Bozio, A., Douenne, T., Fabre, B., et Leroy, C. (2019a). Budget 2019 : quels effets pour les ménages. *Note de l'IPP*, (37).
- Ben Jelloul, M., Bozio, A., Douenne, T., Fabre, B., et Leroy, C. (2019b). Le modèle de micro-simulation taxipp - version 1.1. *Guide méthodologique IPP*.
- Bozio, A., Cottet, S., et Malgouyres, C. (2018). Quels effets attendre de la transformation du cice en réductions de cotisations employeurs? *Note de l'IPP*, (36).
- Bozio, A. et Geoffard, P.-Y. (2017). L'accès des chercheurs aux données administratives. *Rapport du Cnis*.
- Direction technique de l'Agirc-Arrco (2016). Résultats 2016. *Chiffre'AGIRC-ARRCO*. Brochure disponible uniquement en ligne.

Douenne, T. (2018a). Les effets redistributifs de la fiscalité carbone en France. *Note de l'IPP*, (34).

Douenne, T. (2018b). The vertical and horizontal distributive effects of energy taxes : A case study of a French policy. *FAERE Working Paper*, (2018.10).

Fabre, B., Guillouzouic, A., Lallemand, C., Leroy, C., et Malgouyres, C. (2019). Les impacts du budget 2020 sur les ménages et les entreprises. *Présentation - conférence sur l'évaluation du Budget 2020*.

Hill, C. et Legoupil, C. (2018). Taxation et prix du tabac en France et conséquences sur la consommation. *Bull Epidemiol Hebd*, (14-15), p. 309–316.

Ministère des solidarités et de la santé (2017). Rapport d'évaluation de la prime d'activité.

Sénigout, E. et Duval, J. (2016). L'emploi dans la fonction publique en 2014. *Insee Première*, (1586).



L'Institut des politiques publiques (IPP) est développé dans le cadre d'un partenariat scientifique entre PSE-Ecole d'économie de Paris (PSE) et le Centre de Recherche en Economie et Statistique (CREST). L'IPP vise à promouvoir l'analyse et l'évaluation quantitatives des politiques publiques en s'appuyant sur les méthodes les plus récentes de la recherche en économie.

PSE a pour ambition de développer, au plus haut niveau international, la recherche en économie et la diffusion de ses résultats. Elle rassemble une communauté de près de 140 chercheurs et 200 doctorants, et offre des enseignements en Master, École d'été et Executive education à la pointe de la discipline économique. Fondée par le CNRS, l'EHESS, l'ENS, l'École des Ponts-ParisTech, l'INRA, et l'Université Paris 1 Panthéon Sorbonne, PSE associe à son projet des partenaires privés et institutionnels. Désormais solidement installée dans le paysage académique mondial, la fondation décloisonne ce qui doit l'être pour accomplir son ambition d'excellence : elle associe l'université et les grandes écoles, nourrit les échanges entre l'analyse économique et les autres sciences sociales, inscrit la recherche académique dans la société, et appuie les travaux de ses équipes sur de multiples partenariats. www.parisschoolofeconomics.eu



Le CREST est un centre de recherche regroupant des chercheurs de l'ENSAE, de l'ENSAI et du département d'économie de l'Ecole Polytechnique. Centre interdisciplinaire spécialisé en méthodes quantitatives appliquées aux sciences sociales, le CREST est organisé en 4 thématiques: Economie, Statistiques, Finance-Assurance et Sociologie. La culture commune des équipes est celle d'un attachement fort aux méthodes quantitatives, aux données, à la modélisation mathématiques, et d'allers-retours continus entre les modèles théoriques et les preuves empiriques permettant d'analyser des problématiques sociétales et économiques concrètes. <http://crest.science>

