

LE NECESSAIRE ECHANTILLONNAGE DES DONNEES FISCALES : LE CAS DE L'IMPOT SUR LE REVENU

Lorraine AEBERHARDT¹ (*), Ludwin GUENCHI (*) et Rémi LARDELLIER (*)²
(*) DGFIP, Bureau des études statistiques en matière fiscale (GF-3C)

Résumé

Dans le cadre de ses travaux d'évaluation des politiques fiscales, le SSM des finances publiques produit chaque année un échantillon à partir des données exhaustives d'émission d'impôt sur le revenu. La base exhaustive comportant trop d'individus (36 millions de foyers) et les délais de traitement étant restreints, il est en effet impensable de travailler sur l'ensemble de la population. Cet article présente les réflexions et les techniques statistiques testées afin d'actualiser et refondre la méthodologie de constitution de cet échantillon.

Contrairement aux problématiques courantes en théorie des sondages, l'objectif du travail présenté ici n'est pas d'estimer au mieux des grandeurs inconnues. Il s'agit plutôt de réduire le plus fidèlement possible une population dont tous les paramètres sont connus. En plus de la nécessité de réduire les temps de traitements, l'échantillon créé doit également pouvoir être produit rapidement.

L'article détaille les utilisations faites de cet échantillon et donc, les problématiques auxquelles il doit permettre de répondre avec fiabilité. Cette réflexion, illustrée de nombreux travaux passés, permet de sélectionner les variables essentielles dans le renouvellement de la méthodologie d'échantillonnage et pour les travaux à venir. Il est en effet nécessaire de se poser les bonnes questions concernant les déterminants de l'impôt dans un contexte législatif évoluant chaque année.

L'intérêt d'une remise à plat méthodologique réside dans la volonté d'adapter l'échantillon aux nouvelles caractéristiques de l'impôt ainsi qu'aux utilisations qui en sont couramment faites. Surtout, l'idée est de pouvoir profiter des nouveaux outils disponibles pour la production d'échantillons, notamment les méthodes d'échantillonnage équilibré développées par l'Insee (Cube et FastCube).

La dernière partie de l'étude détaille donc la mise en œuvre d'un tirage équilibré avec la macro FastCube. La fiabilité des techniques mises en œuvre est testée et validée tout au long de l'article, point par point, mais également, à la fin, grâce à une comparaison globale de l'ancienne et de la nouvelle méthodologie d'échantillonnage.

¹ lorraine.aeberhardt@dgifp.finances.gouv.fr, ludwin.guenchi@dgifp.finances.gouv.fr, remi.lardellier@dgifp.finances.gouv.fr,

² Les auteurs tiennent à remercier Sébastien Faivre (Insee, chef de la division Sondages au sein de la direction de la Méthodologie et de la coordination statistique et internationale) pour ses précieux conseils. Les auteurs restent évidemment seuls responsables des limites ou inexactitudes de leur travail.

Abstract

The statistical service of the Public Finances General Directorate (DGFIP) evaluates tax policies. As part of this task, a representative sample is built every year from exhaustive income tax data. These data contain all the households concerned with income tax, which represents around 36 million households. Yet, the statistical service often has to simulate fiscal policy projects at a micro level in very short times. As a result, it is not possible to work on the complete data base and a sample needs to be used instead.

This article presents statistical works which have led to updating and rebuilding the sampling method used by the statistical service.

In general, survey sampling methods aim at estimating as precisely as possible unknown variables. On the contrary, in this article, the goal is to reduce as exactly as possible a population for which all parameters are known. Further, the sample needs to be quickly built because the statistical service of DGFIP is asked to produce updated estimations very shortly.

The article describes in details current or past works for which the sample is used or has been used. Moreover, the articles proves that yearly moving legislative framework complicates the income tax determinants definition. This paper draws attention to a few crucial points in building the sample, and shows how to select the main variables of interest.

The goal of rebuilding the sampling method is to take into account new income tax characteristics and new uses of the sample in day-to-day tasks of the statistical service. This is also an opportunity to exploit new statistical sampling methods, in particular balanced sampling by means of the Cube and Fastcube methods developed by INSEE.

As a result, the last part of this article explains in detail how to build a balanced sample using the Fastcube method. Statistical reliability is checked step by step, and finally a global comparison between the old and the new method is proposed.

Mots-clés

Échantillonnage, équilibrage, fiscalité

Introduction

Le bureau GF-3C est depuis 2011 reconnu en tant que service statistique ministériel. Il produit des analyses sur l'ensemble du domaine de la fiscalité, aussi bien en fiscalité des entreprises (TVA, impôt sur les sociétés), fiscalité des particuliers (impôt sur le revenu, droits de mutations, ISF) qu'en fiscalité locale (taxes d'habitation et foncière). L'étude que nous proposons ici porte sur les outils et les données utiles à la production des travaux en matière d'impôt sur le revenu (IR). La section en charge de l'IR a notamment pour missions, d'une part, de participer aux travaux de prévisions de recettes fiscales d'IR (*i.e.* prévision d'un montant global) et, d'autre part, d'estimer le coût de dispositifs fiscaux spécifiques (*évaluations ex ante* ou *ex post*). Pour tous ces travaux, le premier enjeu consiste à concilier précision des résultats et respect de délais contraints.

Pour chaque foyer fiscal français soumis à l'impôt sur le revenu, la base de données de la DGFIP recense l'ensemble des informations déclarées, variables de gestion et variables calculées. Elle est ainsi constituée de plus de 36 millions de foyers et d'environ 3 000 variables. Les données portant sur les revenus de l'année N sont disponibles par vagues (« émissions³ »), de juillet N+1 à février N+2. La richesse d'une telle base est évidente mais elle ne va pas sans difficultés dès qu'il s'agit de mettre en oeuvre des procédures complexes.

Ainsi, en matière de prévision des recettes d'IR ou d'évaluation de dispositifs fiscaux, le bureau GF-3C procède par microsimulation. Un programme retranscrit fidèlement la législation afin de simuler, pour chaque foyer, le calcul de l'impôt. La législation de l'IR étant complexe, ce programme implique de très longs temps de traitement (1h/100 000 foyers en législation 2013, 1h45/100 000 foyers en législation 2014). Dès lors, en vue d'effectuer ces travaux de microsimulation, il est impensable de travailler à partir de la base exhaustive.

Contrairement aux problématiques courantes en théorie des sondages, nous ne cherchons pas à estimer au mieux des grandeurs inconnues, mais à réduire le plus fidèlement possible notre population afin de recalculer l'IR le plus précisément possible et d'en déduire les coûts de dispositifs fiscaux. En plus de la nécessité de réduire les temps de traitements, nos objectifs sont à la fois le maintien d'une bonne précision globale, une reproduction fidèle de la distribution de certaines variables fondamentales dans le calcul de l'impôt, mais également un court délai pour la production de l'échantillon.

Cette problématique n'est pas nouvelle au sein du bureau GF-3C. En 1997, des travaux d'échantillonnage avaient été réalisés, adaptés aux contraintes informatiques de l'époque. Ils ont abouti à la mise en place de deux échantillons, un échantillon dit « léger » (50 000 foyers) et un second dit « lourd » (500 000 foyers). Ces deux échantillons reposaient sur une méthode identique, à savoir un sondage stratifié avec allocation de Bankier et tirage systématique avec tri préalable. Hormis quelques modifications mineures et l'abandon de l'échantillon léger, cette méthode est toujours utilisée pour produire les échantillons d'IR.

Aujourd'hui, l'intérêt d'une remise à plat de cette méthodologie est multiple. Il convient d'abord de l'adapter aux nouvelles caractéristiques de l'IR ainsi qu'aux utilisations devenues courantes de l'échantillon. Dans ce cadre, il est nécessaire de se poser les bonnes questions concernant les déterminants de l'IR dans un contexte législatif évoluant chaque année, en particulier sur la sélection des variables dont nous souhaitons reproduire la distribution. L'idée est surtout de pouvoir profiter des nouveaux outils disponibles pour la production d'échantillons, notamment les méthodes d'échantillonnage équilibré développées par l'Insee (Cube et FastCube). Par ailleurs, contrairement à ce qui était fait précédemment, une production automatisée de l'échantillon permettrait un gain de temps non négligeable lors de la réception des différentes données. Enfin, à long terme, une recherche sur la taille optimale de l'échantillon réduirait les temps de traitement lors des microsimulations.

Dans un premier temps, nous présenterons le fonctionnement général de l'impôt sur le revenu ainsi que les principaux ordres de grandeur des agrégats qui le constituent. Dans un deuxième temps, nous

³ Une émission se constitue au fur et à mesure des traitements des déclarations par les services de la DGFIP.

recenserons les travaux réalisés par le bureau en matière d'IR, en mettant en avant ceux qui nécessitent l'utilisation d'un échantillon. La troisième partie sera l'occasion d'un questionnement sur les critères qu'il convient de reproduire fidèlement pour obtenir un échantillon pertinent au vu des travaux réalisés. Enfin, les quatrième et cinquième parties permettront de détailler, respectivement, l'ancienne et la nouvelle méthode d'échantillonnage.

1. L'impôt sur le revenu : législation et constitution des données du POTE

1.1. Qui est soumis à l'impôt sur le revenu ?

Les foyers tenus de déclarer leurs revenus sont ceux dont la résidence habituelle, le lieu de séjour principal, l'activité professionnelle principale ou les intérêts économiques principaux se situent en France.

Cette obligation concerne également les personnes non imposables⁴. Elles reçoivent ainsi un avis, indispensable pour effectuer certaines démarches, et obtenir le bénéfice d'avantages (exonérations ou allègements de taxe d'habitation, de taxe foncière, de contribution à l'audiovisuel public, avantages sociaux, perception de la prime pour l'emploi).

Les personnes non domiciliées en France doivent aussi souscrire une déclaration de revenus, si elles disposent de revenus de source française (elles sont alors imposées sur ces seuls revenus), ou si elles disposent en France d'une ou de plusieurs habitations. L'imposition de ces personnes ne s'applique que sous réserve des conventions internationales conclues par la France.

Pour les personnes mariées, le foyer fiscal est constitué du contribuable, de son conjoint et des personnes à charge. Les revenus et les charges de tous les membres du foyer fiscal sont pris en compte pour établir une imposition unique au nom du contribuable. L'impôt sur le revenu tient compte de la situation de famille (célibataire, marié, séparé, pacsé, ...), des charges de famille (enfants mineurs à charge, majeurs rattachés,...), et de certaines situations personnelles (invalidité, ancien combattant,...).

Les enfants majeurs (célibataires ou ayant fondé un foyer distinct) doivent souscrire personnellement une déclaration pour les revenus dont ils ont disposé de leur majorité jusqu'au 31 décembre de l'année. Toutefois, ils peuvent demander, sous certaines conditions, le rattachement à l'un ou des parents. L'enfant majeur rattaché n'a pas à déposer de déclaration. Ses revenus sont portés sur la déclaration du ou des parents.

En cas de changement de situation de famille (mariage, PACS, divorce ou séparation), une seule déclaration doit être déposée par les époux ou partenaires de Pacs au titre de l'année du mariage ou de la conclusion du Pacs. Elle doit comporter l'ensemble des revenus dont ils ont disposé pour l'année entière. Sur option irrévocable et uniquement au titre de l'année du mariage ou du Pacs, les époux ou partenaires peuvent opter pour l'imposition distincte de leurs revenus : ils déposent alors une déclaration par personne comportant les revenus dont ils ont chacun disposé l'année entière.

Deux déclarations séparées doivent être déposées par chacun des deux ex-conjoints ou partenaires au titre de l'année du divorce, de la rupture de Pacs ou de la séparation. Chaque déclaration doit comporter les revenus reçus à titre personnel ainsi que la quote part des revenus communs pour l'année entière.

En cas de décès, deux déclarations doivent être souscrites l'année du décès de l'un des conjoints : une au nom des conjoints avec les revenus acquis à la date du décès, l'autre au nom du conjoint survivant pour les revenus disponibles après le décès. L'impôt sur le revenu est calculé en retenant la

⁴

Un foyer est non imposable si son impôt avant prise en compte des crédits d'impôt est supérieur à zéro.

situation et les charges de famille, en principe, au 1er janvier de l'année du décès ou au 31 décembre de l'année du décès s'il y a augmentation des charges.

1.2. Quels sont les revenus à déclarer ?

L'impôt sur le revenu est établi sur la totalité des revenus dont disposent les personnes physiques au cours d'une année déterminée. Les revenus perçus par tous les membres du foyer fiscal doivent être repris sur la déclaration.

Les revenus soumis à l'impôt sur le revenu sont répartis en sept catégories :

- 1) les traitements, salaires, pensions, retraites et rentes
- 2) les revenus des placements financiers
- 3) les plus-values et gains divers (ventes de valeurs mobilières...)
- 4) les bénéfices industriels et commerciaux
- 5) les bénéfices non commerciaux des professions libérales et assimilées
- 6) les bénéfices agricoles
- 7) les revenus fonciers

S'y ajoute une huitième catégorie, constituée des rémunérations de certains dirigeants de sociétés (gérants majoritaires de SARL notamment), dont les règles d'imposition sont toutefois similaires à celles des traitements et salaires.

Certains revenus exonérés ne sont pas à déclarer. Ainsi, les prestations familiales ou les intérêts de livret A sont par exemple totalement exonérés d'impôt sur le revenu. D'autres revenus ne sont que partiellement exonérés, comme les salaires versés aux étudiants de moins de 25 ans (exonération sur option) ou les indemnités journalières AT-MP (exonérées à 50%).

Le tableau ci-dessous, extrait de l'annuaire statistique 2013 de l'impôt sur le revenu 2012 ([7]) détaille les différents revenus catégoriels déclarés par les foyers.

Tableau 1 – Effectifs et montants de revenus catégoriels pour les foyers imposés et non imposés – année 2012 – en millions d'euros

Catégorie de revenus	Foyers imposés		Foyers non imposés	
	Effectif	Montant des revenus déclarés	Effectif	Montant des revenus déclarés
Traitements et salaires	13 816 456	494 762	9 937 789	147 171
Gains sur levée d'options	995	21	122	1
Pensions et rentes	7 092 090	193 599	6 728 863	80 503
B.A. Forfait réel	91 258	329	93 274	336
	175 687	6 299	104 841	1 245
B.I.C. auto-entrepreneurs professionnels et non professionnels	71 764	741	126 710	1 142
micro professionnels	82 684	1 455	166 275	2 154
réel professionnels	334 084	11 896	204 934	2 363
micro non professionnels	282 655	2 134	83 917	591
réel non professionnels	25 352	291	6 789	32
B.N.C. auto-entrepreneurs professionnels et non professionnels	70 243	644	61 170	563
régime spécial professionnel	119 899	1 494	84 588	617
déclaration contrôlée professionnels	425 147	29 268	63 497	1 120
non professionnels régime spécial	37 368	225	18 934	81
non professionnels	13 078	615	2 388	27
Revenus fonciers régime micro	991 602	5 762	486 687	2 081
Revenus fonciers - revenus	1 889 707	26 229	424 445	2 480
- déficits	686 017	11 254	242 658	3 812

R.C.M.	10 453 066	19 132	5 898 981	2 943
Plus et moins-values professionnelles à court terme	471	7	268	1
Plus et moins-values professionnelles à long terme	17 148	1 035	1 748	18
Plus et moins-values (valeurs mobilières)	249 992	4 680	23 454	12

Source : Tableau 202 de l'annuaire statistique 2013 (revenus 2012), DGFIP Bureau GF3-C.

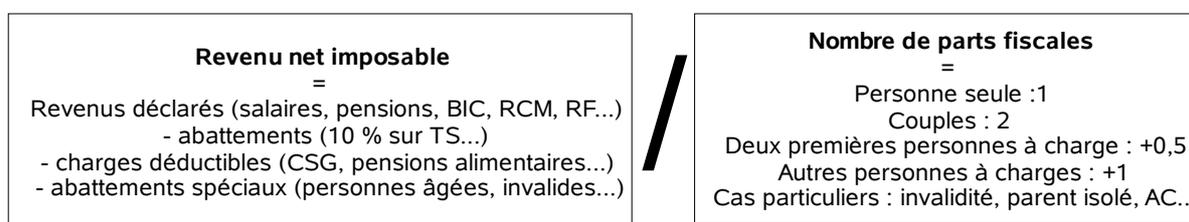
Champ : ensemble des foyers fiscaux.

Note de lecture : Montant en millions d'euros ; les revenus exonérés ne sont pas pris en compte dans ce tableau. Les locations meublées non professionnelles sont réparties entre les BIC non professionnels, micro ou réels.

En montants, les revenus les plus importants sont les traitements et salaires et les pensions et rentes. En effectifs, les traitements et salaires ainsi que les revenus de capitaux mobiliers sont les plus déclarés.

1.3. Calcul de l'impôt sur le revenu

Le calcul de l'impôt repose sur le montant du quotient familial, c'est-à-dire sur le montant des revenus du foyer après prise en compte de la composition familiale. Le schéma ci-dessous synthétise les principales étapes du calcul de l'impôt, à législation 2014.



= Quotient familial imposé par tranche aux taux suivants :



= Impôt au barème progressif, corrigé des montants suivants :

- + plafonnement du quotient familial
- décote
- réductions d'impôt
- + impôts complémentaires à taux forfaitaires
- crédits d'impôt

= Impôt final net, montant négatif, nul ou positif

Source : DGFIP-bureau GF-3C, à partir du formulaire n°204 1 FDC de la brochure pratique 2014.

Encadré n°1 – Dispositifs venant modifier le montant de l'impôt au barème progressif

Plafonnement du quotient familial : le quotient familial correspond au revenu imposable divisé par le nombre de parts fiscales du foyer. Il s'agit donc d'un outil de redistribution horizontale puisqu'à revenu imposable donné, plus le nombre de parts est important, moins l'impôt est élevé. Cependant l'avantage procuré par un grand nombre de parts fiscales est plafonné. En 2014, ce plafonnement est fixé à 1500 euros (dans le cas général) par demi-part au-delà de la ou des parts conjugales. Ainsi,

pour un foyer de trois personnes (un couple marié et une personne à charge, soit 2,5 parts fiscales), la variation d'impôt résultant de la prise en compte ou non de la demi-part associée à la personne à charge ne doit pas excéder 1500 euros. Afin d'effectuer ce plafonnement, une double liquidation de l'impôt est systématiquement réalisée. Elle permet de comparer les impôts avec et sans personnes à charge.

Décote : il s'agit d'un dispositif réduisant automatiquement l'impôt des foyers dont l'impôt après plafonnement du quotient familial n'excède pas 1016 euros (en 2014). Un montant maximal de 508 euros est alors déduit automatiquement de leur impôt.

Réductions d'impôt : certaines dépenses ou investissements donnent droit, pour le contribuable, à une réduction du montant de l'impôt. Ainsi, les dons, les investissements locatifs, les services d'aide à la personne, les dépenses d'accueil dans un EPHAD ou le fait d'avoir des enfants à charge poursuivant leurs études permettent, par exemple, de telles réductions. La réduction de l'impôt se fait dans la limite du montant à payer. Une réduction d'impôt ne permet pas d'avoir un impôt négatif.

Crédits d'impôt : il s'agit d'une créance fiscale qui s'impute sur l'impôt sur le revenu dû, ou fait l'objet d'un remboursement total ou partiel si le montant de l'impôt dû est nul ou inférieur à celui du crédit. Les frais de garde des enfants de moins de six ans, les cotisations syndicales ou les dépenses en faveur de l'amélioration de la qualité environnementale de l'habitation principale ouvrent droit à un crédit d'impôt.

Le calcul de l'impôt sur le revenu repose essentiellement sur la notion de taux marginal. En effet, les taux (progressifs) présentés ci-dessus s'appliquent aux revenus par partie. Par exemple, une personne seule percevant pour tout revenu un salaire annuel imposable de 13 000 euros (après abattement) paiera :

- 0 euro sur la fraction inférieure à 6 011 euros
- 328,9 euros pour la fraction comprise entre 6 011 euros et 11 991 euros (imposée à 5,5%)
- 141,3 euros pour la fraction comprise au-delà de 11 991 euros (imposée à 14%)

À noter que pour un montant total de près de 470 euros, un tel foyer bénéficie du dispositif de la décote en faveur des ménages modestes (cf. encadré n°1), qui réduit son impôt de 273 euros, pour un montant final de 197 euros.

Pour l'ensemble de la population française, le tableau ci-dessous présente, les effectifs de foyers fiscaux et les montants d'impôt associés à chaque tranche marginale d'imposition ([7]).

Tableau 2 – Répartition des foyers et de l'impôt par taux marginal d'imposition

Quotient familial	Taux marginal	Nombre de foyers	Répartition en %	Montant total de l'impôt net ⁵ (en millions d'euros)	Répartition en %
Inférieur à 5 963 euros	0,0 %	8 741 670	23,8	-792	-1,3
Compris entre 5 963 euros et 11 896 euros	5,5 %	8 866 253	24,1	-1 025	-1,6
Compris entre 11 896 euros et 26 420 euros	14,0 %	14 827 094	40,4	18 824	30,2
Compris entre 26 420 euros et 70 830 euros	30,0 %	3 877 237	10,6	26 177	42,1
Compris entre 70 830 euros et 150 000 euros	41,0 %	350 123	1,0	11 216	18,0
Supérieur à 150 000 euros	45,0 %	57 659	0,2	7 841	12,6
Ensemble de la population		36 720 036	100,0	62 242	100,0

Source : Tableau 219 de l'annuaire statistique 2013 (revenus 2012), DGFIP Bureau GF3-C.

Champ : ensemble des foyers fiscaux.

Note de lecture : en 2013, 23,8 % des foyers fiscaux ont déclaré un quotient familial inférieur à 5 963 euros. Ces foyers ont bénéficié, au total, de 792 millions d'euros de restitutions.

⁵ Il s'agit du seul impôt net déclaré sur les déclarations 2042. Il n'est donc pas égal à l'impôt sur le revenu comptabilisé dans les comptes de l'État, qui tient compte d'autres recouvrements, ainsi que de remboursements et dégrèvements, à la suite de contrôles fiscaux ou à contentieux.

Ces résultats montrent qu'un peu moins de 50 % des foyers fiscaux sont imposés au maximum à un taux marginal de 5,5 %. Globalement, ces foyers sont bénéficiaires au titre de l'impôt sur le revenu, puisqu'ils ne paient pas ou peu d'impôt et perçoivent un total de plus de 1 800 millions d'euros, du fait du bénéfice des crédits d'impôt. A l'autre extrémité de l'échelle d'imposition, à peine 60 000 foyers sont imposés à un taux marginal de 45 % : leur impôt payé représente 12,6 % de l'impôt total.

1.4. Le traitement des données d'impôt sur le revenu : de la déclaration n°2042 au POTE

Au printemps de l'année N, chaque foyer fiscal est tenu de déclarer ses revenus de l'année N-1. Depuis les revenus 2005, la déclaration (en ligne ou papier) est préremplie par l'administration fiscale avec les principaux revenus : traitements et salaires, allocations de pré-retraite, de chômage, indemnités journalières de maladie, heures supplémentaires exonérées, pensions et revenus de capitaux mobiliers perçus. Ces informations sont transmises à l'administration fiscale par les organismes payeurs : employeurs, caisses de retraite, Pôle emploi, URSSAF.

Une fois sa déclaration complétée, le foyer envoie sa déclaration à l'administration. Chaque dossier est manuellement validé par un agent et un avis d'imposition sur les revenus est « émis ». Avant d'être agrégées au niveau national, les données sont gérées localement par les établissements de services informatiques de la DGFIP. Ces services initialisent, corrigent et enrichissent le fichier POTE (fichier Permanent des Occurrences de Traitement des Emissions) au fur et à mesure de la disponibilité des « émissions ».

Au total, pour une année de revenus N, six émissions sont généralement nécessaires pour traiter définitivement l'ensemble des dossiers, ce travail débutant à l'été N+1 et se terminant au début de l'année N+2. Les trois premières émissions fournissent près de 99 % du montant des impositions primitives et constituent donc l'essentiel de la base de données. Les trois émissions suivantes sont partagées entre les dernières émissions primitives et les émissions correctives qui modifient, à la marge seulement, le montant global d'impôt émis par l'administration fiscale.

Le POTE contient tous les éléments se rapportant à une déclaration, aussi bien les éléments déclarés que les éléments calculés. La base ne contient qu'une seule ligne par déclaration 2042, quels que soient les revenus et le nombre de traitements effectués au cours des différentes émissions. Chaque déclaration est référencée par l'identifiant FIP (Fichier d'Imposition des Personnes, sur 20 caractères) du premier déclarant du foyer.

Même s'il est utilisé à des fins statistiques par le bureau GF3C, le POTE est avant tout une base de données administratives. En effet, en gestion fiscale, cette base permet tout d'abord d'effectuer des contrôles de vraisemblance avec la situation du contribuable l'année précédente, ainsi que des contrôles de cohérence entre les différentes déclarations qu'un même contribuable aurait pu effectuer au cours d'une année de taxation. Cette base permet en outre d'éditer une déclaration 2042 pour l'année de revenus suivante, et, éventuellement à l'initiative du service, de remettre en cause la seconde déclaration d'un contribuable qui aurait déjà été préalablement taxé suite à une première déclaration.

Pour la constitution de l'échantillon de l'impôt sur le revenu, deux émissions ont été, historiquement, retenues. La deuxième émission (été N) sert généralement à réaliser un échantillon anticipé, utile pour les nombreux travaux de prévisions et simulation de l'automne N. Le second échantillon, définitif, est réalisé à partir de la cinquième émission (décembre N) et constitue la base de l'ensemble des travaux de microsimulation de l'année N+1.

2. Les travaux du SSM des Finances Publiques en matière d'impôt sur le revenu

En matière d'impôt sur le revenu, les travaux du bureau GF-3C peuvent être regroupés en deux catégories : d'une part les travaux qui correspondent à une simple analyse des données présentes dans le POTE et, d'autre part, des travaux, plus lourds, de microsimulation dans le cadre d'évaluations de dispositifs fiscaux. Le bureau GF-3C répond aussi bien à des questions émanant de la sphère publique (parlementaires, cabinets ministériels, autres directions ministérielles, organismes de contrôle) qu'à des demandes de certaines associations ou entreprises de droit privé. Enfin, en tant que SSM, le bureau assure une mission de diffusion de l'information statistique en matière fiscale, à travers notamment la publication en ligne de données synthétiques sur l'impôt sur le revenu et la mise à disposition progressive des bases de données fiscales aux chercheurs.

2.1. Cadre général : les différents travaux

Certains travaux requièrent seulement d'agréger les données départementales disponibles dans le POTE et d'isoler les variables demandées. Si cela n'est pas instantané, ces demandes peuvent toutefois être traitées dans des délais raisonnables et peuvent donc être réalisées à partir des données exhaustives (cf. 2.2). Ainsi, le bureau GF-3C réalise régulièrement des synthèses de suivi des émissions d'impôt (déclarations 2042 nationales, IRCOM, annuaire statistique en ligne, Atlas fiscal) et traite des demandes ponctuelles concernant certains dispositifs isolables dans les données exhaustives (suivi de la PPE ou de certaines réductions et crédits d'impôt etc.).

D'autres demandes, tout aussi nombreuses, ne peuvent toutefois pas être traitées sur les données exhaustives. Il s'agit de tous les travaux nécessitant des simulations de législation. En particulier, les évaluations *ex ante* et *ex post* de nouveaux dispositifs fiscaux sont réalisés de cette façon. La méthode consiste à comparer le niveau global d'impôt en l'absence du dispositif au niveau global d'impôt lorsque le nouveau dispositif est appliqué. La différence entre les deux montants correspond ainsi au coût estimé de la mesure.

La difficulté réside alors dans la durée du calcul puisque, du fait des multiples liquidations qui doivent être réalisées pour obtenir le montant final de l'impôt payé par un foyer, la simulation de l'impôt pour 36 millions de foyers durerait plusieurs semaines. À titre d'exemple, en législation 2014, une simulation d'impôt pour 500 000 foyers est réalisée en 9 heures. C'est donc dans le cadre de ces travaux d'évaluations de mesures fiscales qu'un échantillon s'avère être indispensable.

Deux travaux majeurs du bureau GF-3C relèvent de cette catégorie : d'une part les évaluations de dépenses fiscales *a posteriori* dans le cadre de la rédaction des annexes au projet de loi de Finances de l'année N, et d'autre part, les exercices de prévisions des émissions d'IR pour l'année en cours et à venir (cf. 2.3).

2.2. Demandes traitées sur données exhaustives

2.2.1. Portrait de l'impôt sur le revenu par niveau géographique : l'IRCOM

L'IRCOM (impôt sur le revenu au niveau communal) est un ensemble de tableaux synthétiques décrivant les revenus déclarés et l'impôt émis, pour plusieurs niveaux géographiques du pays ([8]). Ces données sont publiées en ligne pour chaque commune de France, chaque département et chaque région ainsi qu'au niveau national. En plus d'être accessibles au grand public, ces tableaux servent, entre autres, d'indicateurs pour la détermination des dotations aux collectivités locales calculées par le ministère de l'Intérieur.

Ce travail de suivi annuel des données de l'impôt sur le revenu peut être effectué directement sur les données du POTE. En effet, les variables utilisées sont le revenu fiscal de référence (RFR) et l'impôt net, variables centrales du POTE qu'il suffit d'extraire au niveau communal et d'agréger ensuite pour former les différents niveaux géographiques.

2.2.2. Publication de données synthétiques sur l'impôt sur le revenu : l'annuaire statistique et l'atlas fiscal

L'annuaire statistique est une publication en ligne de tableaux synthétiques concernant les principales modalités de calcul et les résultats définitifs des différents impôts ([7]). Ainsi, les données concernant l'impôt sur le revenu, permettent de décrire les montants de revenus et d'impôt ainsi que les effectifs associés à diverses catégories de foyers : foyers imposés ou non imposés, foyers selon leur tranche d'imposition, par tranche de RFR et par nombre de parts. Il s'agit aussi de publier des données macroéconomiques et longitudinales sur l'impôt sur le revenu. C'est ainsi que sont retracés, par tranche d'impôt et depuis 1990, les montants de revenus imposables et d'impôt total émis.

Ces tableaux constituent une simple description des données d'émission composant le POTE. Ce travail est réalisé chaque année (N+2 pour revenus N) et ne nécessite pas de microsimulation. Il est donc réalisé sur la base de données exhaustive.

2.2.3. Estimation de certaines dépenses fiscales : l'exemple de la Prime Pour l'Emploi (PPE)

La prime pour l'emploi (PPE) est un dispositif visant à diminuer l'impôt sur le revenu des foyers dont le revenu fiscal de référence (RFR) n'excède pas 16 251 euros (pour une personne seule). Elle est calculée en fonction du montant déclaré des revenus d'activité (différents taux suivant le niveau des revenus), de la situation familiale (situation matrimoniale, couple mono-actif, parent isolé) et du nombre de personnes à charge (majoration forfaitaire de la prime en fonction du nombre de ces personnes).

Chaque année depuis sa mise en place en 2001, dans le cadre de l'évaluation des dépenses fiscales, le bureau GF-3C fournit une estimation du coût de la PPE. Ce coût est aisément estimable puisque les données du POTE contiennent une variable indiquant le montant de la PPE après imputation du RSA (*i.e.* montant effectivement versé au foyer). De plus, la PPE est le dernier dispositif pris en compte dans le calcul du montant final de l'impôt. Ainsi, sa présence ou son absence n'a pas d'impact sur un autre dispositif intermédiaire dans le calcul de l'IR. Le coût final du dispositif correspond donc simplement à la somme des coûts individuels.

Tableau 3 – Dépenses fiscales de la PPE, extraite du PLF 2015

Numéro de la mesure	Calcul de l'impôt Mesure	(en millions d'euros)		
		2012	2013	2014
110227	Prime pour l'emploi en faveur des contribuables modestes déclarant des revenus d'activité	2 900	2 400	2 210
	<i>Mission et programme :</i> Travail et emploi / Accès et retour à l'emploi (P102)			
	<i>Objectif :</i> Aider les personnes actives aux revenus modestes			
	<i>Bénéficiaires 2012 :</i> 6 325 015 ménages			
	<i>Méthode de chiffrage :</i> Simulation			
	<i>Fiabilité :</i> Très bonne			
	<i>Création / modification :</i> 2001 / 2009			
	<i>Fin du fait générateur :</i> dépense fiscale non-bornée			
	<i>Fin d'incidence budgétaire :</i> dépense fiscale non-bornée			
	<i>Réf. CGI :</i> 200 sexies			

2.3. Demandes nécessitant le recours à un échantillon

Les travaux évoqués jusqu'ici portent soit sur des analyses macroéconomiques de l'impôt, soit sur des dispositifs généraux et bien ciblés dans les données du POTE. Le bureau GF-3C réalise néanmoins de nombreuses autres travaux d'estimations de dispositifs divers. Pour tous les dispositifs non isolables dans le POTE ou dont le calcul est dépendant d'autres dispositifs intermédiaires, la méthode d'estimation consiste en une microsimulation de l'impôt sur le revenu. Compte-tenu de la complexité de la législation de l'IR, il n'est pas possible de microsimuler sur données exhaustives, et c'est pourquoi les microsimulations doivent être réalisées sur un échantillon.

2.3.1. Evaluations de dépenses fiscales a posteriori dans le cadre de la rédaction des annexes au projet de loi de Finances de l'année N

Les dépenses fiscales constituent des dispositions législatives ou réglementaires dont la mise en œuvre entraîne pour l'État une perte de recettes et donc, pour les contribuables, un allègement de leur charge fiscale par rapport à ce qui aurait résulté de l'application de la norme, c'est-à-dire des principes généraux du droit fiscal français.

Ainsi, toute mesure impliquant une perte de recettes pour le budget de l'État n'est pas nécessairement une dépense fiscale : une mesure constitue une « dépense fiscale » lorsqu'elle déroge à une législation de base à laquelle on se réfère et qui constitue la norme. La norme fiscale n'est pas définie de façon intangible et est susceptible d'évoluer en fonction de la législation nationale ou communautaire.

En matière d'impôt sur le revenu, la norme de taxation consiste à soumettre l'ensemble du revenu net global des contribuables au barème progressif, composé de cinq tranches d'imposition à taux marginal (en 2014). Ce revenu net global est constitué des différents revenus catégoriels nets, diminués du montant des charges déductibles du revenu global.

Plusieurs mesures à caractère général sont considérées comme faisant partie intégrante des modalités particulières de calcul de l'impôt et ne constituent donc pas des dépenses fiscales. Il s'agit par exemple de l'abattement de 10% pour frais professionnels, qui constitue une mesure de simplification du législateur envers le redevable, du quotient familial, qui constitue une modalité de calcul d'un impôt familialisé, des mécanismes de décote, des dispositions destinées à éviter une double taxation, des dispositions différant l'imposition dans le temps, des modalités de taxation des revenus à caractère exceptionnel.

L'ensemble des évaluations des dépenses fiscales sont rassemblées dans un unique document budgétaire produit chaque année, le tome II du rapport Voies et Moyens, document annexé au Projet de Loi de Finances (PLF). Pour chaque dépense fiscale, trois évaluations sont présentées dans le rapport N+1, produit en septembre N (date de la finalisation du PLF pour N+1) : la première est une estimation *ex post* de l'impact de la dépense en N-1, la seconde est une estimation « en temps réel » de l'impact de la dépense en N, et la troisième est une prévision de dépense pour l'année N+1. Sont également précisés le nombre de foyers fiscaux concernés par la dépense fiscale, ainsi qu'une estimation de la fiabilité du chiffre (« ordre de grandeur », bonne fiabilité, très bonne fiabilité).

Le bureau GF3C, en matière d'impôt sur le revenu, est sollicité pour évaluer l'impact en N-1 d'un très grand nombre de dépenses fiscales (environ 40) . Il s'agit donc d'une évaluation *ex post*, pour laquelle les données sont disponibles depuis le début de l'année N. Si certaines dépenses fiscales, comme la prime pour l'emploi, sont en lecture presque « directe » dans le POTE, certaines nécessitent d'être microsimulées et l'exercice mobilise ainsi presque à plein temps au début de l'été, la section chargée de l'impôt sur le revenu. Les microsimulations de dépenses fiscales sont faites soit sur échantillon, soit directement à partir du POTE lorsque la dépense fiscale ne concerne qu'une petite partie de la population.

A titre d'exemple, l'abattement de 10% sur le montant des pensions de retraite est microsimulé sur l'échantillon.

2.3.2. Prévisions de recettes d'impôt sur le revenu pour l'année en cours et à venir

2.3.2.1. Description générale de la méthodologie de prévision

Dans le cadre des travaux de prévision des recettes d'impôt sur le revenu, l'évolution de l'impôt est décomposée entre, d'une part, une évolution dite spontanée et, d'autre part, l'impact de nouvelles mesures fiscales sur le produit de l'impôt.

La notion d'évolution spontanée renvoie à l'évolution des recettes liée à la dynamique naturelle de l'assiette de l'impôt c'est-à-dire sa progression économique. Dans le cas des recettes dont l'assiette est fortement corrélée à l'activité économique, cette évolution tient principalement aux hypothèses macroéconomiques associées à la loi de finances.

Concrètement, au sein du bureau GF-3C, l'évolution spontanée de l'impôt est estimée à partir des données de revenus de l'année précédente. Les taux d'évolution reposant sur des hypothèses macroéconomiques sont appliqués à chaque revenu déclaré l'année précédente. La législation fiscale est, elle, laissée inchangée. Avec ces paramètres, un impôt théorique est alors simulé. Par comparaison avec la situation constatée l'année précédente, il est possible d'obtenir une estimation de l'impact sur les recettes d'IR de l'évolution macroéconomique des revenus.

Les prévisions d'IR tiennent en outre compte de l'impact des mesures fiscales sur le produit de l'impôt. En vue des discussions parlementaires, le coût de chaque mesure nouvelle est déjà estimé. Par la suite, afin d'anticiper au plus près les recettes d'IR, de nouvelles estimations peuvent être réalisées avec les paramètres adoptés de manière définitive. Parmi les mesures nouvelles à prendre en compte, il en existe deux catégories :

- celles ne consistant qu'à modifier des paramètres déjà existant de l'impôt (nouvelle tranche d'imposition, modification de seuils, de conditions de revenus etc.),
- celles impliquant la collecte de nouvelles informations auprès des foyers déclarant (nouvelle réduction ou crédit d'impôt, fin de l'exonération de certains revenus etc.).

Seules les mesures nouvelles relevant de la première catégorie peuvent être simulées par anticipation, c'est-à-dire à partir des données déclarées l'année précédente. Les autres mesures nouvelles ne peuvent être évaluées dans le cadre de travaux de microsimulation, du fait de l'absence de données d'assiette dans les bases antérieures.

2.3.2.2. Mesures nouvelles microsimulables

Les mesures nouvelles simulées sur échantillon sont très nombreuses et très diverses, tant la législation évolue régulièrement en matière d'impôt sur le revenu. Par conséquent, tous ces dispositifs ciblent des populations très différentes. Certaines mesures sont générales et concernent tous les foyers déclarant des revenus (indexation du barème). D'autres concernent spécifiquement les foyers à revenus peu élevés (RI exceptionnelle) ou inversement ceux à revenus élevés (nouvelle tranche à 45 %, abaissement du plafond du quotient familial). Il peut également s'agir d'une mesure « catégorielle » : l'imposition d'un type de revenus peut être modifiée, comme cela a récemment été le cas pour les revenus de capitaux mobiliers⁶. Enfin, le critère de simulation peut aussi être la structure du foyer : les personnes seules ayant élevé au moins un enfant ou les foyers avec personnes à charges et plafonnant l'avantage du quotient familial sont des exemples récents de ces mesures ciblées.

Le tableau ci-dessous recense différentes mesures nouvelles qui ont pu être microsimulées au cours des trois dernières années dans le cadre des exercices de prévision. Ces mesures concernent principalement des modifications de seuils ou de taux permettant le calcul de l'impôt ou encore la suppression de certains dispositifs intermédiaires.

⁶ Les revenus de capitaux mobiliers pouvaient être taxés, selon le choix du contribuable, soit au barème soit au prélèvement forfaitaire libératoire (taux proportionnel). Depuis 2013, les dividendes et intérêts sont obligatoirement imposés au barème.

Tableau 4 – Description de mesures nouvelles récentes microsimulées par le bureau GF-3C

Mesure	Coût estimé en millions d'euros	Population concernée	Variable déterminant la mesure
Réduction d'impôt exceptionnelle en faveur des foyers modestes	1 349	Foyers modestes	RFR Parts fiscales Situation matrimoniale.
Indexation du barème	- 600	Ensemble des foyers	Tranches d'imposition revenu imposable
Extinction progressive du bénéfice de la demi-part « personnes seules »	255	Personnes ayant élevé seul un enfant moins de 5 ans	Parts fiscales Situation matrimoniale
Création d'une tranche à 45 %	320	Foyers à quotient familial > 150 000 euros	Revenu imposable Parts fiscales Taux marginal
Barémisation de certains RCM	3 278	Foyers percevant des RCM	Revenus catégoriels
Abaissement du plafond de l'avantage procuré par le quotient familial	1 020	Foyers à revenus élevés et personnes à charge	Revenu imposable Parts fiscales

Source : Voies et Moyens tome 1 des PLF 2013, 2014 et 2015.

2.3.2.3. Mesures nouvelles non simulables

Toutes les mesures inscrites dans les Loi de finances et entrant en vigueur pour l'imposition des revenus N ne sont pas simulables. Ces mesures non simulables sont en général des mesures d'assiette. En effet, il arrive souvent que les données d'assiette sur laquelle repose la nouvelle mesure ne soient pas disponibles dans la base de données des revenus N-1. Il n'est donc pas possible d'estimer *ex ante* l'impact de la mesure. En outre, les données d'assiette relatives à la nouvelle mesure sont souvent par la suite agrégées à d'autres données d'assiette, sans qu'il ne soit possible des les isoler. Dans ce cas, il n'est pas possible non plus d'estimer *ex post* l'impact de la mesure. Pour ce type de dispositif, des estimations sont toutefois réalisées, par d'autres services, à partir de modèles économétriques capables de prendre en compte des hypothèses d'évolution de l'assiette de l'impôt sur le revenu.

Dans certains cas très précis, il est possible toutefois de simuler certaines mesures d'assiette. Il s'agit par exemple de mesures dont les revenus étaient, avant l'entrée en vigueur de la mesure, déclarés à l'administration fiscale mais exonérés : ainsi, même s'ils n'étaient pas taxés, les montants de ces revenus qui étaient déclarés, figuraient bien dans le POTE des années antérieures.

A titre d'exemple s'agissant des mesures d'assiette non simulables, l'article 4 de la loi n°2013-1778 du 29 décembre 2013 de finances initiale pour 2014 prévoyait ne de plus exclure du revenu imposable des salariés les cotisations d'assurance maladie complémentaire collective versées par l'entreprise au profit du salarié. Cette mesure revient donc à élargir l'assiette de l'impôt sur le revenu pour les salariés concernés.

Or ces cotisations étant auparavant exonérées, mais elles n'étaient pas déclarées : par conséquent, les données relatives à ces cotisations ne sont pas disponibles dans le POTE des revenus 2012, derniers revenus auxquels ne s'applique pas la mesure. Ainsi, à partir de ces données 2012, antérieures à la mesure, il est donc impossible de microsimuler ce nouveau dispositif.

De même, une fois la mesure appliquée en 2014, il n'est pas possible de microsimuler précisément l'impact de cette mesure car les cotisations sont désormais agrégées aux salaires déclarées. D'autres méthodes, développées par d'autres services, ont donc été utilisées pour les estimations retenues dans les lois de finances.

◆ Suppression de l'exonération fiscale de la prise en charge par l'employeur d'une partie des cotisations de prévoyance complémentaire. 864 000

3. A la lueur de ces travaux, quels objectifs pour l'échantillon ?

L'échantillon est désormais couramment utilisé, dans un contexte législatif qui évolue chaque année. Compte-tenu de la diversité des travaux effectués et de la complexité de la législation, il doit reproduire fidèlement la distribution d'un maximum de variables.

3.1. Mesurer au mieux l'impôt sur le revenu

La plupart des travaux du bureau GF3C nécessitant le recours à un échantillon poursuivent le même objectif : mesurer au mieux l'impact en impôt sur le revenu de telle ou telle mesure fiscale, afin de prévoir au mieux les encaissements d'impôt sur le revenu et de disposer d'évaluations les plus fiables possible dans le cadre des lois de finances. Il est donc indispensable que la variable « Impôt sur le revenu » du POTE, ainsi que ses déterminants qui influent « mécaniquement » par le biais de la taxation, sur l'impact en recettes d'impôt sur le revenu, soient pris en compte dans la constitution de l'échantillon.

Comme les différentes mesures fiscales sont directement liées à de multiples déterminants de l'impôt, et que pour de nombreuses mesures fiscales, seule une partie de la population est concernée, nous avons choisi de stratifier et d'allouer essentiellement en fonction des déterminants de l'impôt sur le revenu. En revanche, afin de garantir une bonne fiabilité du montant d'impôt sur le revenu en lui-même, la variable « Impôt sur le revenu » fait partie des variables retenues pour l'équilibrage.

3.2. Reproduire fidèlement la distribution des revenus...

Le revenu est, de toute évidence, le principal déterminant de l'impôt à payer. En outre, un grand nombre de mesures simulables font intervenir des seuils de revenus fiscal de référence ou la notion de revenu imposable, ce qui nécessite de disposer de sous-échantillons représentatifs sur chacune des tranches de revenus. Notre objectif principal, dans la constitution d'un échantillon, est donc de concilier représentativité des revenus et représentativité de l'impôt.

Le tableau national de l'IRCOM (ci-dessous, [8]) fait d'ailleurs apparaître l'étalement des revenus fiscaux de référence mais aussi la concentration de l'impôt sur les foyers les plus riches. Ainsi, alors que 99,6 % des foyers ont des RFR inférieurs à 200 000 euros, les 0,4 % de foyers restants ont des RFR compris entre 200 000 euros et 3 millions d'euros et représentent plus de 20% de l'impôt. Ce phénomène de concentration de l'impôt, conjugué à des niveaux de revenus atypiques, n'est pas neutre dans le cadre de la réflexion pour l'élaboration d'un échantillon. Ainsi, si l'on souhaite reproduire au mieux la structure des revenus et de l'impôt, il est alors indispensable de sélectionner exhaustivement les foyers les plus riches.

En outre, dans l'échantillonnage, le revenu est pris en compte à toutes les étapes : il intervient dès la stratification, via le revenu imposable qui constitue l'assiette de la taxation, puis lors de l'allocation, via le revenu fiscal de référence qui sert de critère pour de très nombreuses mesures fiscales, puis enfin lors de la dernière étape de l'échantillonnage, qui permet d'équilibrer à la fois sur le revenu imposable, sur le revenu fiscal de référence et sur les différents revenus catégoriels.

Tableau 5 - Tableau national de l'IRCOM 2013

Tranche de revenu fiscal de référence (en euros)	Nombre de foyers fiscaux	% cumulé	Revenu fiscal de référence des foyers fiscaux	% cumulé	Impôt net (total)	% cumulé
0 à 10 000	8 770 812	23,9%	39 175,1	4,2%	-661,9	-1,1%
10 001 à 12 000	2 186 515	29,8%	24 093,9	6,8%	-393,2	-1,7%
12 001 à 15 000	3 549 589	39,5%	48 258,8	11,9%	-49,8	-1,8%
15 001 à 20 000	5 859 624	55,5%	101 688,4	22,8%	2 468,6	2,2%
20 001 à 30 000	6 556 151	73,3%	161 124,0	40,0%	5 869,7	11,6%
30 001 à 50 000	6 145 060	90,1%	234 562,5	65,0%	13 083,6	32,6%
50 001 à 100 000	2 956 205	98,1%	194 278,0	85,7%	17 662,0	61,0%

100 001 à 200 000	547 812	99,6%	72 092,7	93,4%	11 276,6	79,1%
200 001 à 300 000	84 024	99,8%	20 122,8	95,6%	4 213,7	85,9%
300 001 à 400 000	27 769	99,9%	9 515,2	96,6%	2 136,0	89,3%
400 001 à 500 000	12 864	99,9%	5 716,6	97,2%	1 294,4	91,4%
500 001 à 600 000	6 683	100,0%	3 643,9	97,6%	824,9	92,7%
600 001 à 700 000	4 072	100,0%	2 632,1	97,9%	582,4	93,7%
700 001 à 800 000	2 732	100,0%	2 035,8	98,1%	451,2	94,4%
800 001 à 900 000	1 978	100,0%	1 675,1	98,3%	363,0	95,0%
900 001 à 1 000 000	1 518	100,0%	1 439,6	98,4%	304,8	95,5%
1 000 001 à 2 000 000	4 777	100,0%	6 418,6	99,1%	1 304,4	97,6%
2 000 001 à 3 000 000	1 851	100,0%	8 335,8	100,0%	1 511,5	100,0%
Total	36 720 036	100,0%	936 809,0	100,0%	62 242,0	100,0%

Source : Tableau national de l'IRCOM 2013 (revenus 2012), DGFIP Bureau GF3-C.

Champ : ensemble des foyers fiscaux.

Note de lecture : en 2012, 23,8 % des foyers fiscaux avaient un RFR inférieur à 10 000 euros. Le RFR de ces foyers représentait 4,2 % du RFR national et les restitutions perçues par ces foyers s'élevaient à 661,9 millions d'euros.

3.3. ... et les situations familiales

L'impôt sur le revenu est un impôt « familialisé », dont le calcul est étroitement dépendant de la structure familiale. Le quotient familial constitue ainsi un outil de redistribution horizontale, taxant moins les familles nombreuses, à revenu équivalent ([12]).

Du fait de cette interconnexion entre le montant de l'impôt sur le revenu et la composition familiale, l'estimation de l'impact des mesures fiscales repose donc sur une bonne prise en compte de cette composition.

De plus, certaines mesures, tel le plafonnement du quotient familial relient directement la structure familiale au montant d'impôt à payer. D'autres mesures, comme la demi-part supplémentaire en faveur des personnes ayant élevé seul un enfant pendant moins de cinq ans, ne trouvent à s'appliquer qu'à une sous-population, déterminée justement à partir de la structure familiale.

Enfin, le tableau 215 de l'annuaire statistique montre la variabilité des revenus selon la typologie familiale ([7]). A titre d'exemple, les foyers bénéficiant d'au moins 5 parts fiscales ont plus fréquemment des revenus (RFR) supérieurs à 50 000 euros que les autres foyers.

Tableau 6- Répartition des foyers fiscaux selon le RFR et le nombre de parts fiscales

Nombre de parts fiscales	Tranche de revenu fiscal de référence							Ensemble
	Jusqu'à 10 000 euros	De 10 000 à 15 000 euros	De 15 000 à 20 000 euros	De 20 000 à 30 000 euros	De 30 000 à 50 000 euros	De 50 000 à 100 000 euros	Plus de 100 000 euros	
1	0%	17%	36%	31%	12%	3%	1%	100%
de 1,25 à 1,75	0%	8%	32%	35%	18%	5%	1%	100%
2 (mariés ou pacsés)	0%	0%	3%	26%	44%	22%	4%	100%
2 (autres)	0%	0%	6%	49%	32%	10%	2%	100%
de 2,25 à 2,75	0%	0%	0%	17%	49%	28%	6%	100%
3	1%	0%	0%	5%	46%	38%	10%	100%
de 3,25 à 3,75	0%	0%	0%	3%	49%	37%	10%	100%

4	1%	0%	0%	0%	35%	45%	19%	100%
de 4,25 à 4,75	0%	0%	0%	0%	32%	47%	20%	100%
5	1%	0%	0%	0%	19%	46%	33%	100%
5,25 et plus	1%	0%	0%	0%	7%	48%	42%	100%

Source : Tableau 215 de l'annuaire statistique 2013 (revenus 2012), DGFIP Bureau GF3-C.

Champ : ensemble des foyers fiscaux.

Note de lecture : en 2012, 48 % des foyers fiscaux ayant au moins 5,25 parts fiscales avaient un RFR compris entre 50 000 et 100 000 euros.

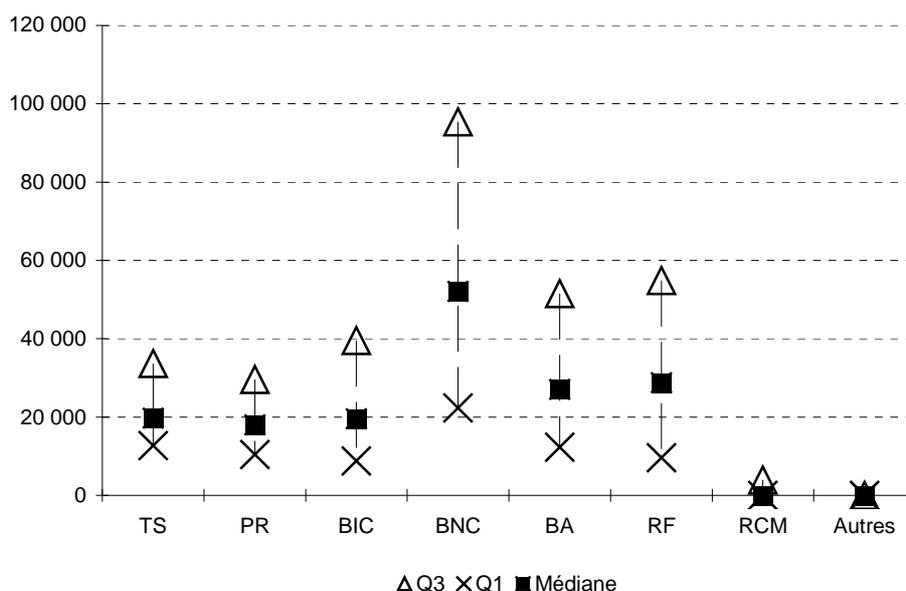
Dans ces conditions, il est indispensable que l'échantillon retrace le plus fidèlement possible la structure familiale de la population. A cet effet, elle a été directement intégrée dans la stratification, et la situation familiale fait l'objet d'un équilibrage.

3.4. Contrôler la répartition des catégories de revenus

Le mode de calcul de l'impôt (cf. 1.3) est ainsi fait que, à montants de revenus nets équivalents, le type de revenus (salaires, pensions, BIC, BNC, bénéfices agricoles, revenus de capitaux mobiliers ou revenus fonciers) est neutre sur le montant d'impôt final. En cela, la catégorie de revenu dominant du foyer n'est pas un déterminant de l'IR. Cependant, le fait qu'un foyer fiscal possède principalement des revenus de telle ou telle catégorie, aux dépens d'autres types de revenus, correspond implicitement à un comportement économique particulier (choix de diversifier ou non ses revenus, en faveur d'une catégorie) et un comportement fiscal spécifique (recours aux réductions et crédits d'impôt) qui peuvent influencer d'autres paramètres de calcul de l'impôt.

En outre, comme le montre le graphique ci-dessous, la distribution des revenus n'est pas la même suivant la catégorie de revenu dominant⁷ du foyer. A titre d'exemple, les foyers recevant principalement des bénéfices non commerciaux (BNC) ont un RFR médian largement supérieur à celui des foyers recevant principalement des pensions. Contrôler la répartition des revenus dominants lors de l'échantillonnage assurerait alors indirectement une bonne distribution des revenus globaux.

Tableau 7- Distribution des RFR selon la catégorie de revenu dominant



Source : POTE des revenus 2013 à la cinquième émission, DGFIP Bureau GF3-C.

⁷ La notion de revenu dominant renvoie à un type de revenus, supérieur à tous les autres revenus perçus par le foyer et constituant plus d'un tiers de ses ressources.

Champ : ensemble des foyers fiscaux.

Note de lecture : en 2013, le RFR médian des foyers dont le revenu dominant étaient les traitements et salaires (TS) s'élevait à environ 20 000 euros.

Enfin, il est important de contrôler la répartition des différents types de revenus du fait de la possibilité d'avoir à simuler certaines mesures catégorielles. Ainsi, pour l'impôt 2014 sur les revenus 2013, une mesure spécifique aux revenus de capitaux mobiliers a dû être simulée. Ces revenus concernant près de 16 millions de foyers fiscaux (*cf.* tableau 1), il était évidemment impossible de microsimuler la mesure sur la base de données exhaustives.

Pour toutes ces raisons, mais sans que cela soit le premier paramètre à contrôler, il convient de tenir compte des catégories de revenus au cours de l'échantillonnage.

3.5. Prendre en compte les dispositifs fiscaux modulant l'IR à revenu égal

Le tableau ci-après (tableaux 205, 206, 207, 208, 209 et 210 de l'annuaire statistique 2013, [7]) fait apparaître certains résultats dont il faut tenir compte dans le cadre de la constitution d'un échantillon du POTE. Ainsi, à revenu donné (RFR ou revenu imposable) le montant d'impôt peut varier. Ce phénomène résulte de la diversité des dispositifs fiscaux existants. Il ne suffit pas de connaître les revenus d'un foyer pour produire une bonne estimation de son impôt. Le montant payé dépend en effet de plusieurs facteurs. Ce tableau détaille donc les principaux dispositifs pris en compte dans le calcul de l'impôt et permettant de réduire le montant d'impôt final. Il peut s'agir de réductions d'impôt (diminution de l'impôt dans la limite d'un montant nul), de crédits d'impôt (réduction sans limite) ou de charges déductibles (réduction du revenu imposable).

Tableau 8 – Effectifs et montants des principaux dispositifs pouvant réduire l'impôt sur le revenu

Dispositifs	Effectif	Montant (en millions d'euros)
Crédits d'impôt		
Salariés à domicile	1 555 692	2 041
Qualité environnementale de l'habitation principale	795 184	668
Intérêts sur habitation principale	1 947 851	1 684
Frais de garde	1 813 669	1 125
PPE après imputation du RSA	5 871 178	2 353
Total des crédits d'impôt	X	9 629
Charges déductibles		
Pensions alimentaires enfants majeurs	3 150 950 880	3 151
Autres pensions alimentaires	4 894 999 698	4 895
Déficits antérieurs	1 322 102 997	1 322
Epargne retraite	2 038 698 817	2 039
CSG déductible	2 132 767 137	2 133
Total des charges	X	14 404
Réductions d'impôt		
Dons aux personnes en difficulté	2 016 695	396
Dons aux œuvres	4 574 054	2 208
Salarié à domicile	2 314 242	5 471
Hébergement de long séjour	404 367	6 731
Souscriptions au capital de PME	63 159	1 497
Investissements locatifs neufs Scellier	223 629	18 511
Total des réductions d'impôt	X	41 174

Source : annuaire statistique 2013 sur les revenus 2012, DGFIP Bureau GF3-C.

Champ : ensemble des foyers fiscaux.

Note de lecture : en 2013, plus d'1,5 million de foyers fiscaux ont bénéficié du crédit d'impôt pour emploi d'un salarié à domicile ; le montant total de ce crédit d'impôt s'est élevé à près de 2 milliards d'euros.

Ces dispositifs sont essentiels dans la constitution d'un échantillon pour deux raisons. D'une part, les travaux de microsimulations réalisés peuvent porter sur des aménagement de l'un de ces dispositifs. D'autre part, à revenu imposable donné, ces dispositifs viennent surtout introduire de l'hétérogénéité

dans le montant final d'impôt. Notre objectif étant de concilier représentativité des revenus et représentativité de l'impôt, il est crucial de contrôler également ces dispositifs.

4. L'échantillon LOU, méthodologie d'élaboration de 1997 à 2014

La méthode de constitution de l'échantillon en vigueur jusqu'en 2014 avait été élaborée en 1997 par J. Harnois ([11]) alors stagiaire Ensae au sein du bureau. Étant donné les ressources informatiques disponibles à l'époque, il avait été choisi de réaliser deux échantillons, l'un de 50 000 foyers dit « échantillon léger », l'autre de 500 000 foyers, dit « échantillon lourd » (d'où le nom de LOU associé à cette base de données).

4.1. Rappel sur la méthode d'allocation de Bankier

Cet échantillon reposait sur une allocation optimale dite de Bankier (ou power allocation, allocation optimale avec objectifs locaux, cf. [2], [1], [13]), elle-même dérivée de l'allocation optimale de Neyman. Dans un sondage stratifié, l'allocation de Neyman est celle qui minimise la variance de l'estimateur de la moyenne de Y, dans le cas d'un sondage aléatoire simple dans chacune des strates, pour un budget donné ou une taille d'échantillon fixée. La méthode de Neyman ne permet toutefois de contrôler qu'une seule variable d'intérêt. La méthode de Bankier permet de limiter les variances au niveau des regroupements d'analyse car l'allocation se fait sur la base de deux variables. L'allocation d'une strate est alors calculée à partir de la formule suivante :

$$n_h = \frac{(X_h)^a \cdot \frac{S_h}{\bar{Y}_h}}{\sum_{h=1}^H (X_h)^a \cdot \frac{S_h}{\bar{Y}_h}} \cdot n$$

Où n_h est l'allocation optimale de la strate h , X_h une variable auxiliaire corrélée à la variable d'intérêt, a un paramètre, S_h la dispersion au sein de la strate, \bar{Y}_h la moyenne de la variable d'intérêt au sein de la strate, H le nombre de strates, n la taille souhaitée de l'échantillon.

L'originalité de cette allocation réside dans le coefficient $(X_h)^a$ qui permet de plus ou moins (selon a) tenir compte de la variable auxiliaire.

4.2. Application de la méthode de Bankier dans le cadre de la constitution de l'ancien échantillon

4.2.1. Stratification

La méthode de Bankier est une méthode pour les sondages stratifiés. La première étape consiste donc à stratifier l'ensemble des foyers fiscaux. Pour cela, trois variables du Pote sont retenues :

- la condition d'imposition : foyer imposable ou non imposable ;
- le montant du revenu imposable augmenté des plus-values imposées au taux proportionnel et des revenus de capitaux mobiliers imposés au barème (nouveau 2014) ;
- la catégorie de revenu dominant : un revenu est dit « dominant » lorsque celui-ci est le plus élevé de tous les revenus catégoriels nets du foyer fiscal et représente au moins le tiers de la somme des revenus catégoriels nets. Si l'une des conditions n'est pas vérifiée, on parle d'absence de revenu dominant.

Le tableau ci-dessous présente les seuils retenus pour stratifier la population suivant le montant des revenus imposables (augmentés des plus-values). Ces montants n'ont jamais été révisés depuis le passage à l'euro et il est difficile de savoir à quelle logique correspond ce découpage.

Tableau 9 – Strates de revenus retenues dans l'ancienne méthode d'échantillonnage

Tranche de revenus	Contribuables imposables	Répartition des foyers en %	Contribuables non imposables	Répartition des foyers en %
1	7 500 €	0,3	2 500 €	20,66
2	10 000 €	0,09	5 000 €	8
3	12 500 €	1,55	7 500 €	12,53
4	15 000 €	9,17	10 000 €	15,3
5	19 000 €	15,25	12 500 €	16,01
6	22 500 €	11,14	15 000 €	7,95
7	27 500 €	13,07	17 500 €	7,25
8	32 500 €	11,16	20 000 €	4,67
9	42 500 €	15,26	22 500 €	2,44
10	61 500 €	13,27	25 000 €	1,45
11	80 000 €	4,67	35 000 €	2,51
12	205 000 €	4,52	44 000 €	0,53
13	Plus de 205 000 €	0,57	Plus de 44 000 €	0,69

Source : POTE 2012 à la cinquième émission, DGFIP Bureau GF3-C.

Champ : ensemble des foyers fiscaux.

Note de lecture : dans l'ancienne stratification des revenus, 20,66 % des foyers non imposables composaient la première tranche.

Il apparaît ainsi que les tailles de strates sont très hétérogènes, alors même que les seuils retenus sont arbitraires. Ainsi la première strate des foyers non imposables en contient 20,66 % alors que chacune des six dernières strates ne contient que moins de 5% de la population.

Une fois les strates élaborées, il est possible de calculer l'allocation optimale de chaque strate. La taille approximative souhaitée pour l'échantillon est de 500 000 foyers fiscaux. Les allocations calculées n'étant pas des entiers, l'arrondi entraîne toujours une taille supérieure à 500 000.

4.2.2. Allocation de Bankier

La variable auxiliaire prise en compte pour la constitution de l'échantillon est le montant des charges du foyer fiscal. Les charges sont constituées des pensions alimentaires, des frais d'accueil d'une personne, de l'épargne retraite et des réparations payées par les nus-proprétaires : ces postes réduisent tous le revenu net global et créent donc des différences à revenu initial identique.

A noter que contrairement à ce qui est indiqué dans la formule théorique présentée dans la première partie, nous prenons $X_h = n_h Y_h$, c'est-à-dire que les charges sont à la fois considérées comme la variable auxiliaire et la variable d'intérêt. Par défaut, la valeur du paramètre a de Bankier est fixée à 0,5.

Enfin, lors de la réalisation de l'échantillon, il arrive que certaines strates aient un montant total de charges nul. De ce fait le calcul d'une allocation selon la méthode de Bankier est impossible, puisque le dénominateur est nul. Pour rendre le calcul possible, nous calculons des charges théoriques à partir d'une version antérieure de cette même strate.

Pour certaines strates, la méthode de Bankier peut conduire à allouer un nombre d'unités échantillonnées supérieur à la population existante. Dans ce cas, la procédure utilisée limite l'allocation au nombre de foyers fiscaux existants : 1 200 foyers sont concernés par ce tirage exhaustif. Par ailleurs, dans le souci de bien représenter les foyers atypiques et fortement contributeurs à l'impôt, les foyers imposables et non imposables appartenant à la treizième tranche de revenus sont tous retenus dans l'échantillon : 200 000 foyers sont concernés par ce tirage exhaustif. En fin de compte, plus de 40 % des foyers de l'échantillon ne représentent qu'eux-mêmes.

4.3. Contrôle par tri préalable et tirage systématique

Afin de réaliser un tirage systématique mais aussi afin de mieux contrôler le tirage, un tri préalable est réalisé. Les variables utilisées pour ce tri sont les suivantes :

- strate croisée avec le nombre de parts fiscales : ce découpage correspond aux « urnes » du tirage
- code de la direction des services fiscaux ; cette variable permet d'assurer un minimum de représentativité géographique
- montant des réductions d'impôt (tri décroissant)

Le tirage systématique suit l'algorithme suivant, sur chacune des strates :

Poids = Effectif Pote / Effectif échantillon

Si début de l'urne alors

b = 0;

Si poids ≥ 2 alors a = poids / 2

Sinon a = poids

Fin Si

b = b + 1

Si b = Ent(a) alors

Sélectionner ce foyer

a = a + poids

Fin Si

Le tirage revient donc à sélectionner tous les individus à intervalle du poids associé à la strate. Seul le premier pas est différent puisqu'en cas de strate non retenue exhaustivement (poids ≥ 2), le poids est divisé par deux. Chaque urne étant triée par montant décroissant des réductions d'impôt, la première étape de l'algorithme permet donc d'éviter « d'oublier » les individus à réductions d'impôt élevées lors du tirage. Cela évite *in fine*, de surestimer l'impôt. A noter qu'un tirage systématique est généralement aléatoire. Dans notre cas, alors qu'il avait été conçu pour être initialisé aléatoirement, l'algorithme utilisé durant les dernières années débutait par la sélection arbitraire d'un foyer.

Le tri selon le montant des réductions d'impôt, tout comme la prise en compte de l'hétérogénéité des charges dans l'allocation de Bankier, ont le même objectif : à revenus identiques, il convient de tenir compte des facteurs qui peuvent faire varier l'impôt final (les charges réduisent les revenus imposables et les réductions d'impôt... réduisent l'impôt).

Sur ce point, le programme de tirage n'est pas à jour puisque depuis la fin des années 1990 les crédits d'impôts sont de plus en plus nombreux dans la législation. Une amélioration serait d'intégrer les crédits d'impôt dans le tri préalable, au même titre que les réductions.

4.4. Validations

Les poids réels (effectifs des strates dans l'échantillon tiré après l'étape de tri/ effectifs des strates du POTE) sont comparés aux poids théoriques (effectifs des strates dans l'échantillon déterminés à partir de l'allocation de Bankier/ effectifs des strates du POTE) afin de s'assurer que ceux-ci ne sont pas trop éloignés l'un de l'autre. Les poids théoriques correspondent à l'inverse du taux de sondage alors que les poids réels sont modifiés du fait du pas de tirage qui est nécessairement un entier.

Une comparaison est pratiquée entre les données agrégées du POTE (données de l'Etat 1507 M au niveau national) et ces mêmes agrégats reconstitués à partir de l'échantillon. Si la différence relative ($[\text{montant de l'agrégat de l'échantillon} / \text{montant de l'agrégat 1507}] - 1$) de chaque agrégat n'est pas comprise dans un intervalle – arbitraire – de +/-30% et que le nombre de foyers est important, cela signifie que l'échantillon n'est pas « représentatif » pour l'agrégat étudié.

5. De LOU à FELIN⁸ : vers un échantillon actualisé et équilibré

L'amélioration de l'échantillonnage passe par deux étapes. Dans un premier temps, il convient de mettre à jour les seuils servant à la stratification des foyers mais également de questionner la pertinence des variables utilisées. Dans un deuxième temps, l'objectif est de remplacer le tirage systématique par un tirage équilibré. Toutes ces modifications ont pour but d'améliorer la précision et la représentativité de l'échantillon sur les variables déterminantes de l'impôt. L'ensemble de ces travaux d'amélioration a été réalisé à partir des données de la cinquième émission de l'impôt 2013 (revenus 2012).

5.1. Optimisation de la stratification

Pour mener à bien la refonte de la stratification de cet échantillonnage, deux options étaient envisageables : stratifier le POTE selon des méthodes d'analyse de données ou bien maintenir une stratification empirique. Après avoir exploré un temps la possibilité d'utiliser des méthodes de classification, nous avons finalement abandonné cette voie. En effet, ces méthodes présentent plusieurs inconvénients pour nos travaux : elles sont difficilement utilisables sur nos données vu leur volumétrie, elles auraient nécessité d'être exécutées à chaque fois qu'un nouveau jeu de données alimentait le POTE, et enfin, les seuils de stratification en résultant auraient parfois été difficilement interprétables. En revanche, une stratification empirique a le mérite d'être facilement réalisable. Surtout, loin d'être hasardeuse, elle s'appuie sur l'excellente connaissance des mécanismes de l'impôt sur le revenu que peuvent avoir les agents du SSM. Nous proposons donc ci-après une refonte de la stratification selon une méthode empirique.

5.1.1. D'imposables à imposés

Le premier point qui mérite d'être amélioré dans la réalisation de l'échantillon LOU concerne la notion d'« imposable », utilisée lors de la stratification. Un foyer imposable est un foyer dont le montant de l'impôt avant prise en compte des crédits d'impôt est supérieur à zéro euro. Cette notion, très souvent citée, est *in fine* mal connue et mal comprise. En effet, le montant de l'impôt avant prise en compte des crédits d'impôt est un résultat intermédiaire rarement diffusé.

Il existe en outre une confusion courante entre cette notion d'imposable, complexe, et la notion de foyer « imposé », beaucoup plus simple, puisqu'elle désigne les foyers fiscaux dont l'impôt final net à payer est strictement supérieur au seuil de recouvrement de 12 euros. En d'autres termes, les foyers imposés sont ceux qui paient de l'impôt et les foyers non imposés sont ceux qui ne paient pas d'impôt ou qui bénéficient d'une restitution fiscale.

Outre les difficultés liées à la compréhension de la notion d'imposable, et qui ne facilitent pas la compréhension des règles de détermination de la stratification, l'utilisation de cette notion d'imposable pose des problèmes strictement méthodologiques. En effet, à niveau donné d'impôt avant crédits d'impôt, l'impôt final net ne sera pas le même suivant l'existence ou non de crédits. Dans ces conditions, l'exploitation de la notion d'imposable vient ajouter de l'hétérogénéité dans les strates ainsi dichotomisées.

Encadré n°2 – Analyse de la variance intra-strate

Comme nous cherchons en priorité à mesurer le niveau d'impôt sur le revenu de façon fiable, l'objectif lors de la construction de l'échantillon, est de minimiser la variance de l'impôt sur le revenu dans l'échantillon obtenu. En conséquence, nous cherchons également, lors de l'élaboration des strates, à optimiser la stratification au regard de la décomposition de la variance de l'impôt sur le revenu sur l'ensemble de la population.

Pour rappel, dans un sondage stratifié avec tirage aléatoire simple dans chaque strate, l'égalité suivante est vérifiée :

⁸ Fichier Echantillonné de L'Impôt sur le reveNu

$$s_y^2 = \sum_{h=1}^H \left(\frac{n_h}{n} (\bar{y}(x_h) - \bar{y})^2 \right) + \sum_{h=1}^H \left(\frac{n_h}{n} s_y^2(x_h) \right)$$

i.e. la variance totale est égale à la somme pondérée des écarts des moyennes par strate à la moyenne globale (variance inter-classes) et de la somme pondérée des variances par strate (variance intra-classe). Une stratification est optimale lorsqu'elle minimise l'hétérogénéité de chaque classe (variance intra-classe) et maximise les contrastes entre strates (variance inter-classes).

Même si – nous le verrons plus loin – nous n'utilisons pas *in fine* un sondage stratifié avec tirage aléatoire simple dans chacune des strates, le principe de cette optimisation, qui repose sur la minimisation de la variance intra-classe et sur la maximisation de la variance inter-classe, peut nous servir de point de repère pour la construction de notre échantillon. Un tirage équilibré étant plus optimal qu'un tirage aléatoire, les résultats produits constitueront des majorants.

Cette intuition est confirmée par le tableau 9, qui compare l'homogénéité des deux classes créées, soit avec la notion d'imposable, soit avec celle d'imposé. La différence est minime mais le critère d'imposé permet d'obtenir des classes plus homogènes du point de vue du montant d'IR.

Tableau 10 – Comparaison de l'homogénéité des découpages suivant le statut d'imposition

	Imposés / Non imposés	Imposables / Non imposables
Variance intra-classe	98,10%	98,54%
Variance inter-classes	1,90%	1,46%
Variance totale	100,00%	100,00%

Source : POTE 2012-5.

Champ : ensemble des foyers.

Note de lecture : L'hétérogénéité de l'impôt sur le revenu, interne aux strates d'imposés/non imposés représente 98,1 % de l'hétérogénéité de la base entière. Avec la stratification imposables/non imposables, ce taux était de 98,54 %. La stratification imposés / non imposés est donc préférable car regroupe des populations plus homogènes du point de vue de l'IR.

Pour toutes ces raisons et en vue d'uniformiser les notions diffusées par le bureau GF-3C qui privilégie désormais une communication sur la notion d'imposés, nous abandonnons la notion de foyer imposable au profit de celle de foyer imposé.

5.1.2. Actualisation et automatiser des seuils de revenus

Les strates de revenus utilisées dans l'ancienne méthode d'échantillonnage correspondent à un découpage en fonction du revenu imposable augmenté des plus-values mobilières à taux proportionnel, découpage pour lequel il est difficile de savoir à quelle logique il correspond. En outre, les montants n'ont jamais été révisés depuis le passage à l'euro, et les strates ainsi formées ne retracent plus fidèlement la distribution de la population selon le revenu.

Par conséquent, nous optons désormais pour une stratification par déciles de revenus, actualisés chaque année. Le passage de seuils de revenus arbitraires à des quantiles permet d'automatiser l'actualisation du découpage, tout en conservant des regroupements pertinents pour l'analyse de la distribution des revenus. Reste alors à déterminer selon quel revenu stratifier.

5.1.3. Prise en compte de la composition du foyer

Nous prévoyons, lors du tirage, d'abandonner la méthode de tri préalable au tirage, au profit de méthodes permettant d'équilibrer un grand nombre de variables dans l'échantillon (cf. 5.3).

Le nombre de parts fiscales du foyer n'était exploité, dans l'ancienne méthodologie, que pour trier les foyers fiscaux avant le tirage : cette notion de parts fiscales du foyer n'intervenait ni dans l'étape de stratification, ni dans l'étape d'allocation. Par conséquent, abandonner dans la nouvelle méthodologie le tri préalable au tirage, revient à supprimer le contrôle du nombre de parts fiscales du foyer dans la

procédure d'échantillonnage. Or il est indispensable de conserver ce contrôle, dans la mesure où, à revenu imposable donné, l'impôt sera différent en fonction du nombre de parts fiscales. En outre, certaines mesures fiscales que nous devons simuler dépendent directement du nombre de parts du foyer fiscal.

C'est pourquoi nous avons étudié deux options pour réintroduire, dès la stratification de l'échantillon, un contrôle sur le nombre de parts du foyer fiscal.

5.1.3.1. Option 1 : remplacement du revenu imposable par le quotient familial

La première méthode consisterait à remplacer la variable de revenus utilisée jusque-là (RIPV, i.e. revenu imposable augmenté des plus-values) par une autre variable de revenu qui prendrait directement en compte le nombre de parts du foyer fiscal.

L'objectif, dans cette démarche, est de ne pas introduire une nouvelle variable dans la stratification existante – à savoir le nombre de parts du foyer fiscal – afin de ne pas augmenter le nombre de strates. En effet, nous le verrons plus loin, un trop grand nombre de strates complexifie le tirage et l'équilibrage. Dans le cadre d'un tirage équilibré, disposer d'effectifs plus grands au sein de chaque strate facilite la convergence vers les totaux demandés.

Afin d'étudier également la nécessité de prendre en compte ou non les plus-values à taux proportionnelles, trois approches sont envisagées ici pour remplacer la variable de revenus actuelle :

- option 1A = (Revenu imposable + Plus-values à taux proportionnel) / Nombre de parts
- option 1B = Quotient familial = (Revenu imposable) / Nombre de parts
- option 1C = Quotient familial + Plus-values à taux proportionnel

Contrairement aux deux autres, l'option 1B ne permet pas de concilier la prise en compte du nombre de parts fiscales et la prise en compte des plus-values imposées à taux proportionnel.

Le tableau 11 ci-dessous présente la part des variances intra-classe de l'IR dans la variance totale par décile de revenus. L'option 1C est celle qui fournit les strates les plus homogènes du point de vue de l'impôt. Ce résultat est logique car l'option 1C correspond à une définition des revenus quasiment équivalente à celle retenue pour définir l'assiette de l'IR.

Tableau 11 – Comparaison de l'homogénéité des options de prise en compte du quotient familial

Stratification	Part de la variance intra-classe dans la variance totale
Actuelle	93,4 %
Option 1A	92,9 %
Option 1B	92,9 %
Option 1C	92,8 %

Source : POTE 2012-5.

Champ : ensemble des foyers stratifiés selon le revenu (suivant l'option).

Note de lecture : L'hétérogénéité interne aux strates de revenu actuelles (en termes d'impôt) représente 93,4 % de l'hétérogénéité de la base entière.

5.1.3.2. Option 2 : stratifier suivant le nombre de parts

La seconde méthode consiste à ajouter le nombre de parts fiscales comme variable de stratification. Par rapport à la méthode précédente, celle-ci a l'avantage de ne pas complexifier la signification de la variable de revenu. En effet, un montant donné de quotient familial peut dissimuler des revenus différents, des charges déductibles différentes ou situations familiales différentes. En ajoutant le croisement avec une variable catégorielle de parts fiscales, il est possible de conserver l'homogénéité du revenu imposable, tout en différenciant les situations familiales.

Cependant, afin de ne pas multiplier les strates, nous testons cette stratification selon le nombre de parts en abandonnant la stratification selon la catégorie de revenu dominant (CRD).

Le schéma ci-dessous décrit les variables et les modalités utilisées pour cette stratification.

Revenu imposable augmenté des plus-values		Nombre de parts fiscales		Statut d'imposition	
D1		1		Imposé	= 120 strates
D2		1,25 – 1,5 – 1,75			
D3			2		
D4		2,25 – 2,5 – 2,75			
D5			3 et plus		
D6					
D7					
D8					
D9					
P95					
P99					
P100					

A partir de ces croisements, 120 strates sont théoriquement construites. En réalité, du fait de la concentration de foyers non imposés en bas de l'échelle des revenus, seulement 110 strates sont créées : les premier et deuxième déciles de foyers non imposés ne peuvent être dissociés car tous les foyers de ces déciles présentent un revenu imposable nul.

5.1.4. Comparaison globale des stratifications envisagées

Il s'agit maintenant de comparer les options 1C et 2 afin de savoir quelle approche est la plus pertinente pour stratifier la base.

5.1.4.1. Degré d'homogénéité des strates créées

Le tableau 12 ci-dessous détaille la part de variance intra-classe pour les trois méthodes de stratification étudiées (actuelle, option 1 et option 2). Chaque analyse est réalisée à la fois sur la population entière (POTE à la cinquième émission) et la seule population sondée. En effet, les foyers présentant les plus hauts revenus (dernier centile) sont tirés exhaustivement, et il est peu pertinent de les prendre en compte dans l'analyse de la variance intra-classe.

Les strates de hauts revenus concentrent environ 94 % de la variance intra-classe totale (pour tous les échantillons testés). Cette hétérogénéité importante vient d'ailleurs conforter le choix de l'exhaustivité de la dernière strate de revenus.

Tableau 12 – Comparaison de l'hétérogénéité des options de stratification envisagées

Variable d'intérêt	Champ	Ancienne stratification	Option 1 C	Option 2	Option 2 avec RFR (pour validation)
Revenu imposable	Sondés	5,62 %	31,76 %	4,4 %	6,3 %
	Tous	20,87 %	44,8 %	30,9 %	32,2 %
RFR	Sondés	31,86 %	58,27 %	42,3 %	5,3 %
	Tous	67,14 %	75,88 %	72,6 %	68,5 %
Impôt sur le revenu	Sondés	33,72 %	29,38 %	19,7 %	78,4 %
	Tous	74,0 %	76,8 %	76,4 %	81,6 %

Source : POTE 2012-5.

Champ : ensemble des foyers.

Note de lecture : Les strates résultant de l'ancienne stratification contiennent 20,87 % de l'hétérogénéité des revenus imposables de l'ensemble de la base. Sur la seule population qui ne sera pas retenue exhaustivement dans l'échantillon, cette hétérogénéité diminue à 5,62 %.

Il ressort de cette comparaison que les deux options de stratification permettent d'améliorer l'homogénéité des strates du point de vue de l'impôt sur le revenu pour la sous-population sondée ; le gain est notamment important dans le cas de l'option 2 (de 34 % de variance intra-strates actuellement à 20 %). Concernant le revenu imposable, seule l'option 2 permet d'améliorer

l'homogénéité au sein des strates : de 5,6 % de variance actuellement à 4,4 %. Concernant le revenu fiscal de référence, les strates sont moins homogènes avec les deux options. Néanmoins l'option 2 est celle qui dégrade le moins l'homogénéité, selon le RFR, des strates par rapport à l'ancienne stratification.

Du point de vue de l'analyse de l'homogénéité des classes (suivant l'IR, le RFR et le RIMP), il semble donc que l'option 2 soit préférable à la méthode actuelle et à l'option 1. L'option 2 a également pour avantage de reposer sur un nombre restreint de strates (120 contre 234 actuellement et 216 avec l'option 1). Ce critère est important dans le cadre d'un tirage équilibré puisqu'il implique a priori des effectifs plus grands au sein de chaque strate, ce qui facilite la convergence vers les totaux demandés.

Nous avons également vérifié qu'il n'était pas plus pertinent, dans la stratification avec l'option 2, d'exploiter le RFR au lieu du revenu imposable augmenté des plus-values. Il apparaît que le revenu imposable augmenté des plus-values est bien la variable de stratification qui permet de minimiser la variance intra-classe, pour les trois variables d'intérêt étudiées.

5.2. Choix de la variable d'allocation

Au-delà de la qualité de la stratification, nous étudions également le niveau de précision associé à la stratification choisie et à une allocation de Neyman. Le calcul de cette précision est donné par la variance de la moyenne estimée de la variable d'intérêt, soit :

$$V(\hat{y}) = \sum_{h=1}^H \left(\frac{N_h}{N} \right)^2 \times \left(1 - \frac{n_h}{N_h} \right) \times \frac{S_h^2}{n_h}$$

Où N_h/N est le poids observé de la strate h dans l'ensemble de la population, n_h/N_h , le taux de sondage dans la strate h et S_h^2 , la variance observée de la variable d'intérêt dans la strate h . Pour rappel, l'allocation optimale de Neyman est déterminée suivant la formule suivante :

$$n_h = n \times \frac{N_h \cdot S_h}{\sum_{h=1}^H N_h \cdot S_h}$$

Étant donné le délai de réalisation d'un équilibrage sur nos données, nous calculons des précisions à partir des tailles allouées théoriquement, à partir de l'allocation de Neyman. Les précisions calculées sont donc des précisions théoriques avant équilibrage. Par ailleurs la variance utilisée pour le calcul de la précision n'est pas une variance estimée mais la variance réelle calculée sur le POTE complet (nous disposons des données exhaustives). Enfin, ces calculs reposant sur un tirage aléatoire simple, les résultats présentés ci-après ne constituent que des majorants des précisions réelles.

Une des questions centrales de ce travail est de savoir quelle est notre variable d'intérêt. Or il est impossible de trancher entre les variables d'IR, de RIMP et de RFR : suivant les travaux réalisés (modification ou non de la législation notamment) la variable privilégiée n'est pas la même. A ce titre, les différentes étapes de l'échantillonnage (stratification, allocation et équilibrage) sont l'occasion de parvenir à contrôler ces différentes variables en même temps.

Il s'agit donc de calculer la précision de l'échantillonnage pour différentes valeurs S_h . Quatre variables ont été testées : les montants de charges déductibles, d'impôt, de revenu imposable et de revenu fiscal de référence (RFR). Dans le tableau ci-dessous, le montant des charges, jusque-là utilisé comme variable d'intérêt de l'allocation sert de référence pour comparer les trois autres allocations.

Tableau 13 – Comparaison des précisions associées à différentes variables d'allocation

Variable d'allocation	Evolution de la variance (précision)
-----------------------	--------------------------------------

	Impôt sur le revenu	Revenu imposable	Revenu fiscal de référence
Charges	réf.	réf.	réf.
Impôt sur le revenu	-64%	-50%	-60%
Revenu imposable	-52%	-72%	-56%
Revenu fiscal de référence	-52%	-61%	-68%

Source : POTE 2012-5.

Champ : ensemble des foyers.

Note de lecture : le passage d'une allocation fonction des charges à une allocation fonction de l'IR permet de diminuer de 64 % la variance de l'estimateur de l'IR.

Il convient de noter que du fait de l'étude de la variance de trois variables différentes, les précisions présentées pour une variable d'allocation donnée peuvent constituer des majorants. Par exemple, pour la ligne « impôt sur le revenu », l'allocation est fonction de l'IR donc le -64% traduit l'amélioration de la précision de l'IR. En revanche, le -50% associé au revenu imposable (RIMP) ne porte pas sur l'évolution de la précision réelle mais sur l'évolution d'un majorant de la précision, car l'allocation n'a pas été réalisée suivant le RIMP.

A la lecture du tableau, il ressort que l'allocation en fonction de l'impôt sur le revenu permet une forte amélioration de la précision de l'impôt, mais dans des proportions qui restent limitées. Ainsi, l'allocation en fonction de l'IR ne permet un gain que de -64% sur la précision de l'IR. *A contrario*, concernant la précision du revenu imposable, l'allocation en fonction de cette même variable permet un gain très important de -72 %. De même, concernant la précision du RFR, la stratégie d'allocation en fonction du RFR est performante, de -68%.

En revanche, l'allocation en fonction de l'impôt sur le revenu ne permet qu'une amélioration moyenne de la précision des deux variables de revenu (-50 % pour le revenu imposable, -60 % pour le RFR). Inversement, les deux allocations en fonction du revenu (RIMP ou RFR) sont équivalentes du point de vue de l'amélioration de la précision de l'impôt (-52%).

En fin de compte, si l'allocation fonction de l'IR semble moins intéressante que les deux autres, aucune des deux allocations fonction du revenu ne semble plus optimale que l'autre. Alors que le revenu imposable est déjà pris en compte lors de la stratification, le RFR n'est jusque-là pas contrôlé. Afin de s'assurer d'une certaine représentativité de l'échantillon du point de vue du RFR, nous décidons d'effectuer l'allocation en fonction de cette variable de revenu.

Enfin, le tableau ci-dessous détaille la précision de l'ancien échantillon et de l'échantillon Option 2 suivit d'une allocation en fonction du RFR. Le gain de précision est, pour chacune des variables testées, très important pour la nouvelle stratégie adoptée.

Tableau 14 – Comparaison des niveaux de précision entre l'ancienne et la nouvelle méthode d'échantillonnage

Stratégie	Variance de l'estimateur et % d'amélioration		
	Impôt sur le revenu	Revenu imposable	Revenu fiscal de référence
<i>Ancienne stratification et allocation de Bankier fonction des charges</i>	16,5	72	584
	-84,2%	-75,0%	-54,6%
<i>Option 2 avec allocation de Neyman fonction du RFR</i>	2,6	18	265

Source : POTE 2012-5.

Champ : ensemble des foyers.

Note de lecture : le passage de l'ancienne à la nouvelle stratification permet de réduire la variance de l'estimateur de l'impôt sur le revenu de 84,2 %, faisant passer la précision de 16,5 à 2,6.

5.3. Mise en place d'un tirage équilibré avec FastCube

Le tirage de l'échantillon constitue la dernière étape de la procédure d'échantillonnage. Afin d'améliorer l'échantillon existant, nous testons ici le passage d'un tirage systématique à un tirage équilibré. Si le gain théorique est évident du point de vue de la qualité de l'échantillon produit, la mise en place d'un tel tirage s'avère en revanche être complexe étant donné nos ressources informatiques.

5.3.1. Pourquoi équilibrer ?

Les deux premières étapes de l'échantillonnage, la stratification et l'allocation de Neyman, ont assuré une certaine représentativité du point de vue de l'impôt sur le revenu, du revenu imposable et du revenu fiscal de référence. Mais de nombreux dispositifs ne reposent pas uniquement sur ces trois variables.

D'après l'analyse faite précédemment sur les mesures nouvelles microsimulées, certaines variables, autres que ces dernières, doivent également être représentées fidèlement. Par exemple, la mesure relative à la barémisation des revenus de capitaux mobiliers, ne peut être estimée correctement sans une reproduction fidèle des revenus de capitaux mobiliers. Plus généralement, les mesures fiscales concentrées sur une catégorie spécifique de revenus sont fréquentes, d'où l'importance de construire un échantillon représentatif au point de vue des revenus catégoriels (huit catégories). C'est notamment pour cette raison que, lors de la mise en place de l'ancienne méthode d'échantillonnage, le contrôle de ces variables de revenus catégoriels était effectué dès l'étape de stratification.

D'autres variables autres que les revenus catégoriels mais tout aussi essentielles pour les estimations n'ont pas été prises en compte dans la nouvelle méthodologie développée dans cet article, ni dans l'étape de stratification, ni dans l'étape d'allocation. La méthode de tirage doit donc permettre de rendre l'échantillon représentatif également sur toutes ces variables, et c'est pourquoi la méthodologie retenue est celle d'un tirage équilibré.

Contrairement au tirage systématique, le tirage équilibré permet en effet un contrôle sur un nombre plus vaste de variables, en particulier des variables de revenus catégoriels, qu'il est désormais possible de contrôler à ce stade de l'échantillonnage

La méthode du tirage équilibré consiste à sélectionner un nombre défini de foyers tel que l'estimation des totaux des variables souhaitées, connues *a priori* sur l'ensemble de la population, soit parfaite. A la suite des travaux de Deville et Tillé ([5], [6]), l'Insee a développé une macro-SAS permettant d'effectuer un tirage équilibré, la macro CUBE.

5.3.2. Faisabilité d'un équilibrage de l'échantillon IR : Cube ou FastCube ?

La macro CUBE est un programme SAS mis à disposition par l'Insee permettant d'assurer le quasi équilibrage d'un échantillon ([14]). En effet, en pratique il est difficile d'obtenir un équilibrage parfait étant donné l'étendue des variables à équilibrer et l'inégalité des probabilités d'inclusion. Cet algorithme doit trouver un équilibre entre les probabilités d'inclusion, les équations d'équilibrages et l'aspect aléatoire du plan de sondage. Pour trouver l'équilibre, cet algorithme réalise des tirages sans remise visant à minimiser l'éloignement des contraintes lorsqu'il n'est pas possible de satisfaire exactement les équations d'équilibrage, et à respecter au mieux les probabilités d'inclusion.

Étant donnée la taille importante de notre population à échantillonner (36 millions de foyers) et le court délai de production, il est impossible d'exécuter CUBE sur l'ensemble de notre base en l'état avec les capacités informatiques dont nous disposons⁹.

Reprenant les principes de CUBE, la macro FastCube développée par Chauvet et Tillé ([3], [4]) permet, comme son nom l'indique, un gain non négligeable dans le temps d'exécution d'un équilibrage (environ 3000 fois plus rapide !). Ce gain de temps s'explique par le fait que l'algorithme

⁹ Notamment, le serveur sur lequel est stocké l'ensemble des données ne permet pas de travailler avec SAS-IML. Ce module n'est installé que sur les équipements individuels, ce qui implique de réaliser l'équilibrage en local, avec moins de puissance informatique.

ne cherche plus l'échantillon optimal dans l'ensemble de la base à chaque étape, mais seulement dans un sous-vecteur de dimension $p+1$ (p étant le nombre de variables d'équilibrage).

Afin de réaliser le tirage sur des populations à probabilité d'inclusion identiques, nous n'appliquons pas FastCube à l'ensemble de la base stratifiée, mais à chaque strate l'une après l'autre. Les variables sélectionnées seront donc équilibrées à la fois au niveau des strates et au niveau global. Lors de la phase d'atterrissage (sélection des dernières unités), nous utilisons l'option 1 du paramétrage de la macro. Celle-ci consiste à libérer les contraintes une à une (en débutant par la dernière variable inscrite dans le paramétrage) tant que l'algorithme n'a pas convergé. Si celui-ci converge le foyer est sélectionné. Les deux autres options d'atterrissage n'ont pas pu être testées car elles nécessitent la procédure LP d'optimisation, non installée avec les versions de SAS présentes au bureau GF-3C.

5.3.3. Sélection des variables d'équilibrage

La reproduction fidèle de la répartition des revenus, des situations familiales, des charges déductibles et de celles ouvrant droit à une réduction ou crédit d'impôt, des taux marginaux et enfin de notre variable « d'intérêt », l'impôt sur le revenu, est importante pour répondre aux différentes questions auxquelles nous sommes confrontés.

Encadré 3 - Variables d'équilibrage testées

Ces variables sont testées car leur rôle est important dans la détermination de l'impôt. Il s'agit de :

- la constante ; elle permet de fixer la taille de l'échantillon. En effet, la taille de l'échantillon issu du tirage peut être différente de celle prévue. Cette variable permettra de respecter les proportions des effectifs au sein de chaque strate.
- le revenu fiscal de référence (RFR) vise à estimer les capacités contributives réelles des redevables dans la mesure où il tient compte notamment des revenus ou plus-values taxés au taux proportionnel. La variable retenue est le montant du revenu fiscal de référence
- le revenu net imposable (RIMP) sert à déterminer le quotient familial et, partant, le montant de l'impôt au barème. La variable retenue est le montant du revenu net imposable.
- les revenus catégoriels regroupent les revenus des foyers, classés en plusieurs catégories pour le calcul de l'impôt sur le revenu : traitements et salaires (TS), pensions et rentes (PR), revenus de capitaux mobiliers (RCM), revenus fonciers (RF), bénéfiques agricoles (BA), bénéfiques industriels et commerciaux (BIC) et bénéfiques non commerciaux (BNC). Les variables retenues sont exprimées en montant.
- les réductions d'impôt (RI). La variable retenue est le montant.
- le crédit d'impôt (CI). La variable retenue est le montant.
- les charges déductibles (CH) du revenu brut global tel que les pensions alimentaires. La variable retenue est le montant.
- la situation familiale du foyer fiscal (FAM) se divise en cinq catégories, à savoir marié, pacsé, célibataire, divorcé et veuf. Une indicatrice est créée pour chaque catégorie, valant 1 si le foyer appartient à la catégorie, 0 sinon.
- le taux marginal d'imposition (TX) ; il se décompose en six tranches (*cf. supra*). Une indicatrice est créée pour chaque tranche, celle-ci valant 1 si le foyer relève de cette tranche.
- Enfin, l'impôt sur le revenu final net (IR). La variable retenue est le montant de l'impôt sur le revenu net.

L'objectif de cette partie est de trouver le meilleur échantillon en utilisant ces variables d'équilibrage. Plusieurs échantillons seront donc constitués afin de les comparer. Dix tirages sont réalisés par échantillon afin d'éviter le biais de l'aléa du tirage. Réaliser dix tirages est un bon compromis entre la durée de traitement de la macro FastCUBE (5 heures par tirage) et la robustesse de l'échantillon.

Le tableau ci-dessous présente l'emboîtement des différents échantillons testés. Le premier échantillon sera équilibré sur un nombre restreint de variables alors que le dernier sera équilibré avec toutes les variables citées précédemment.

Tableau 15 - Variables d'équilibrage retenues pour chaque échantillon testé

Échantillon	Variable d'équilibrage																
	Cst	IR	RFR	RIMP	RI	CI	CH	FAM	BIC	BNC	BA	RCM	RF	DIV	TS	PR	TX
1	X	X	X	X	X	X	X	X									
2	X	X	X	X	X	X	X	X	X	X	X	X					
3	X	X	X	X	X	X	X	X	X	X	X	X	X	X			
4	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
5	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Note de lecture : le premier échantillon est équilibré sur 8 variables, soit 12 totaux (la variable FAM se compose de 5 modalités). Pour le deuxième échantillon, 4 variables sont ajoutées.

L'échantillon 1 est équilibré à partir des variables d'équilibrage suivantes : la constante, l'impôt sur le revenu, le revenu fiscal de référence, le revenu imposable net, les réductions et crédit d'impôt et les charges. Cet ensemble de variables d'équilibrage constitue la base sur lesquels tous les autres échantillons vont être équilibrés. Les revenus professionnels (BIC, BA et BNC) et les revenus de capitaux mobiliers (RCM) ont été ajoutés pour équilibrer l'échantillon 2. L'échantillon 3 reprend les variables d'équilibrage de l'échantillon 2 avec en plus les revenus du patrimoine (RF, DIV). L'échantillon 4 repart de l'échantillon 3 avec en plus les revenus d'activité (TS, PR) et enfin l'échantillon 5 possède en plus de l'échantillon 4, les taux marginaux.

Les résultats des tableaux 16 et 17 correspondent à la moyenne des écarts, en valeur absolue, des montants de la variable étudiée entre chaque tirage d'échantillon et le pote. Soit :

$$Ecart = \frac{1}{10} \times \sum_{i=1}^{10} \sqrt{\left(\frac{x_i - X}{X}\right)^2}$$

où x_i est le total mesuré sur l'échantillon i et X le vrai total constaté sur le POTE. Le tableau représente les écarts relatifs moyens entre les totaux du POTE et ceux des différents échantillons testés. Les variables sont en majorité les variables servant à l'équilibrage.

Tableau 16 – Ecart relatif moyen entre les totaux du POTE et ceux des différents échantillons testés

Variables	Échantillons					POTE (total)	
	AM	1	2	3	4		5
Impôt sur le revenu	0,06%	0,03%	0,02%	0,03%	0,06%	0,02%	62 085 649
Revenu fiscal de référence	0,08%	0,05%	0,02%	0,04%	0,03%	0,03%	934 469 431
Revenu imposable	0,04%	0,00%	0,00%	0,00%	0,00%	0,00%	882 209 950
Traitements et salaires	0,07%	0,08%	0,06%	0,09%	0,01%	0,01%	571 374 987
Pensions et rentes	0,18%	0,25%	0,14%	0,25%	0,02%	0,01%	247 121 269
Bénéfices industriels et commerciaux	0,11%	1,05%	0,29%	0,34%	0,30%	0,37%	17 796 762
Bénéfices non commerciaux	0,13%	0,35%	0,12%	0,12%	0,21%	0,10%	33 546 329
Bénéfices agricoles	0,78%	1,63%	0,38%	0,33%	0,37%	0,26%	8 403 672
Revenus fonciers	0,98%	0,25%	0,49%	0,50%	0,07%	0,06%	33 002 892
Revenus de capitaux mobiliers	0,01%	0,56%	0,20%	0,19%	0,23%	0,21%	15 249 008
Divers	0,64%	0,33%	0,47%	0,36%	0,25%	0,28%	10 712 898
Crédits d'impôt	0,48%	0,15%	0,15%	0,08%	0,12%	0,14%	9 765 640
Réductions d'impôt	0,50%	0,11%	0,11%	0,10%	0,09%	0,12%	6 096 953
Charges	1,43%	0,77%	0,55%	0,68%	0,47%	0,95%	10 945 402
Nombre	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	36 549 910

Source : POTE 2012-5.

Champ : échantillon ancienne méthode (AM), 5 échantillons nouvelle méthode (NM) et POTE complet.

Note de lecture : l'écart moyen (sur 10 jeux de données) entre le total d'IR de l'échantillon n°1 et le total du POTE est de 0,03%. Pour l'échantillon testé n°2, cet écart est réduit à 0,02%.

Les échantillons élaborés avec la nouvelle méthodologie sont, globalement, plus représentatifs selon les variables d'équilibrage qu'avec l'ancienne méthodologie. Deux revenus sont, tout de même, mieux représentés avec l'ancienne méthodologie, il s'agit des bénéfices industriels et commerciaux et surtout les revenus de capitaux mobiliers, l'écart calculé est de 0,01 % avec l'ancienne méthode, et au minimum de 0,19 % pour l'échantillon 3.

Tous les échantillons sont très bien représentés du point de vue de l'impôt sur le revenu (entre 0,02 % et 0,06%), du revenu fiscal de référence (entre 0,03 % et 0,05%) et du revenu imposable (0,00%). L'ajout des variables d'équilibrage relatives aux revenus catégoriels permet d'améliorer très nettement la qualité des échantillons.

Le choix de l'échantillon s'oriente ici vers l'échantillon 4 qui semble être le plus représentatif. Malgré le constat d'une légère dégradation de certaines variables déjà très bien représentées (IR, BIC, BNC), celui-ci apporte une plus grande représentativité sur les autres variables.

Les revenus d'activité, TS et PR, qui sont les revenus les plus déclarés, ont un écart moyen très faible par rapport au POTE, de 0,01 % pour les TS et 0,02 % pour les PR. Ces écarts se sont améliorés de plus de 0,6 point pour les TS et 0,16 point pour les PR par rapport à l'ancien échantillon.

En annexe 1, ces indicateurs sont décomposés entre foyers imposés et non imposés. Cette analyse en sous-groupes d'imposition n'infirme pas les résultats précédents.

Le tableau suivant représente les écarts relatifs entre les effectifs de foyers par situation familiale du POTE et des différentes échantillons testés.

Tableau 17 - Ecart relatif entre les effectifs de foyers par situation familiale du POTE et des différents échantillons testés

Situation familiale	Échantillons					POTE (total)	
	AM	1	2	3	4		5
Célibataire	0,05%	0,00%	0,00%	0,00%	0,00%	0,01%	14 381 570
Divorcé	0,72%	-0,01%	-0,01%	-0,03%	-0,01%	-0,02%	5 255 485
Marié	0,01%	0,00%	0,01%	0,00%	0,00%	0,00%	12 098 758
Pacsés	0,14%	0,01%	0,00%	0,06%	0,06%	0,01%	807 499
Veufs	0,71%	0,01%	0,01%	0,01%	0,00%	0,00%	4 006 598

Source : POTE 2012-5.

Champ : échantillon ancienne méthode (AM), 5 échantillons nouvelle méthode (NM) et POTE complet.

Note de lecture : l'écart moyen (sur 10 jeux de données) entre l'effectif réel de foyers célibataires et l'effectif de l'échantillon n°1 est de 0%.

Ce tableau montre que l'ancien échantillon est moins représentatif que chacun des échantillons élaborés avec la nouvelle méthodologie du point de vue de la situation matrimoniale. L'équilibrage (avec la variable FAM) améliore très nettement la représentation des structures familiales. Les échantillons 1 et 2 semblent plus représentatifs du fait d'un nombre de variables d'équilibrage restreint.

Pour le tableau 18, l'indicateur est la somme des moyennes des écarts pondérés en valeur absolue soit :

$$Ecart_{strates} = \sum_{h=1}^{120} \left[\frac{1}{10} \times \sum_{i=1}^{10} \sqrt{\left(\frac{x_{ih} - X_h}{X_h} \right)^2} \times \left(\frac{n_h}{N} \right) \right]$$

où x_{ih} est le total mesuré sur la strate h de l'échantillon i et X_h le vrai total constaté sur la strate h du POTE. L'intérêt de ce dernier tableau est de s'assurer que la représentativité globale observée précédemment va de pair avec la représentativité au sein de chaque strate.

Tableau 18 - Écart relatif moyen entre les totaux par strates, du POTE et des différents échantillons testés

Variables d'équilibrage	Échantillons					
	AM	1	2	3	4	5
Impôt sur le revenu	5,95%	0,55%	0,52%	0,64%	0,54%	0,71%
Revenu fiscal de référence	0,91%	0,30%	0,33%	0,30%	0,41%	0,32%
Revenu imposable	0,25%	0,01%	0,01%	0,01%	0,01%	0,01%
Traitements et salaires	5,20%	1,54%	1,39%	1,39%	0,26%	0,33%
Pensions et rentes	6,73%	2,53%	2,50%	2,35%	0,19%	0,21%
Bénéfices industriels et commerciaux	18,46%	15,61%	4,11%	4,97%	5,22%	4,53%
Bénéfices non commerciaux	25,72%	25,89%	8,55%	8,62%	10,03%	9,90%
Bénéfices agricoles	33,54%	26,74%	9,21%	8,98%	8,25%	8,05%
Revenus fonciers	11,18%	10,75%	11,38%	10,60%	2,64%	2,78%
Revenus de capitaux mobiliers	9,02%	10,25%	2,99%	2,66%	2,98%	3,07%
Divers	21,00%	25,44%	27,69%	10,28%	10,28%	10,74%
Crédits d'impôt	3,98%	0,63%	0,58%	0,69%	0,57%	0,74%
Réductions d'impôt	21,08%	52,02%	46,75%	46,09%	68,71%	48,74%
Charges	6,94%	4,60%	4,68%	4,89%	4,52%	5,34%

Source : POTE 2012-5.

Champ : échantillon ancienne méthode (AM), 5 échantillons nouvelle méthode (NM) et POTE complet.

Note de lecture : l'écart moyen (sur 10 jeux de données) calculé par strate entre le total d'IR de l'échantillon n°1 et le total du POTE est de 0,55%. Pour l'échantillon testé n°2, cet écart est réduit à 0,52%.

Il ressort de cette analyse par strate que l'échantillon 4 permet aussi d'être plus représentatif au sein de chaque strate. Malgré un nombre de variables d'équilibrage beaucoup plus important que pour les

trois premiers échantillons, les pertes en précisions ne concernent que quelques variables et en faible proportion (RFR, BIC, RI ...). Surtout, pour des variables essentielles, comme les traitements et salaires et les pensions et rentes, le gain en précision semble majeur.

Pour des raisons de faibles performances informatiques (trop de variables d'équilibrage) et de non convergence de CUBE sur certaines strates, l'échantillon 5 bien que plus représentatif sur certaines variables par rapport à l'échantillon 4 sera donc exclu de notre étude.

En fin de compte, tous les échantillons issus de la nouvelle méthodologie sont plus précis sur les variables utiles à nos travaux par rapport à l'échantillon issu de l'ancienne méthodologie qui ne bénéficiait d'aucun équilibrage.

5.4. Validation *ex post* de l'échantillonnage

Cette partie vise à apprécier la qualité du nouvel échantillon produit. L'annexe 1 présente le tableau de validation habituel réalisé en comparant les agrégats de l'Etat 1507 M national à son équivalent produit à partir de l'échantillon (*cf.* annexe 2). Surtout, il importe ici de comparer le nouvel échantillon, à la fois aux données réelles du POTE, mais aussi, aux données résultant de l'ancienne procédure d'échantillonnage.

5.4.1. Validation par microsimulation : instauration d'une tranche d'imposition marginale à 45 %

Une première approche pour tester la validité du nouvel échantillon consiste à microsimuler des mesures fiscales sur chacun des échantillons (ancienne et nouvelle méthode) et si possible à comparer avec le coût réel de la mesure lorsqu'il est disponible. Cet exercice est réalisable pour deux mesures récentes.

La première mesure concerne l'imposition 2013 des revenus 2012. Il s'agit de la mesure créant une nouvelle tranche marginale d'imposition au taux de 45 % pour les foyers dont le quotient familial dépasse 150 000 euros. Cette mesure ne concernant que peu de foyers (moins de 100 000 foyers), il est possible de la microsimuler sur les données du POTE et donc d'obtenir le gain exact de cette mesure pour l'Etat. Sur le POTE 2012 à la cinquième émission des revenus, cette mesure est estimée à 320 millions d'euros (*cf.* tableau 19).

La même microsimulation a été réalisée sur l'ancien et le nouvel échantillon. Alors que l'ancien échantillon surestime de deux millions d'euros le gain, le nouvel échantillon ne le surestime que d'un million.

Tableau 19 - Coûts réel et estimés de la création d'une tranche d'imposition à 45 % en 2013

Base de données	Gain pour l'Etat de la nouvelle tranche (en millions d'euros)
POTE 2012-5	344
LOU 2012-5	346
FELIN 2012-5	345

Source : POTE, LOU et FÉLIN 2012-5.

Champ : ensemble de la population.

Note de lecture : estimé à partir des données exhaustives du POTE, le gain pour l'Etat lié à la création d'une nouvelle tranche d'imposition marginale s'élève à 344 millions d'euros.

A première vue, ces écarts sont dérisoires. Pour autant ils montrent une réelle amélioration de l'échantillonnage, pour deux raisons. D'une part, la population concernée par cette mesure est majoritairement la population retenue exhaustivement dans les échantillons (à 97% environ). Le simple fait d'avoir modifié le mode de sélection d'une faible part de la population cible de cette mesure a amélioré l'estimation. Il est donc semblable que l'amélioration soit plus grande encore pour une mesure qui concernerait une population qui ne serait pas - comme dans le cas de cette mesure fiscale - presque sondée exhaustivement. Surtout, si l'impact ici du nouvel échantillonnage est faible (+0,4 % d'écart contre +0,6 % auparavant), il est évidemment à rapporter au montant en jeu : plus de 300

millions d'euros. Dès lors qu'il s'agirait d'un exercice de prévision du montant global d'impôt sur le revenu¹⁰, une diminution de 0,2 point impliquerait une plus grande précision à hauteur de 160 millions d'euros, ce qui est loin d'être négligeable.

5.4.2. Analyse par bootstrap de la précision de l'échantillon

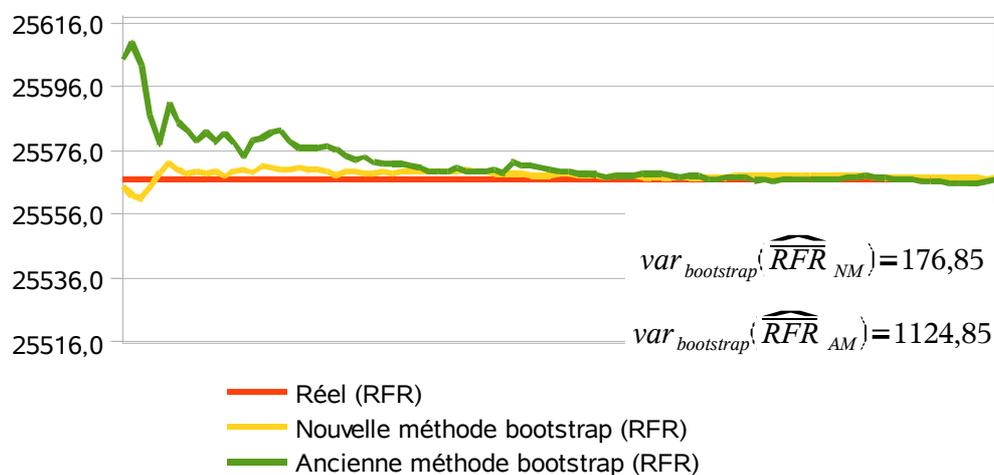
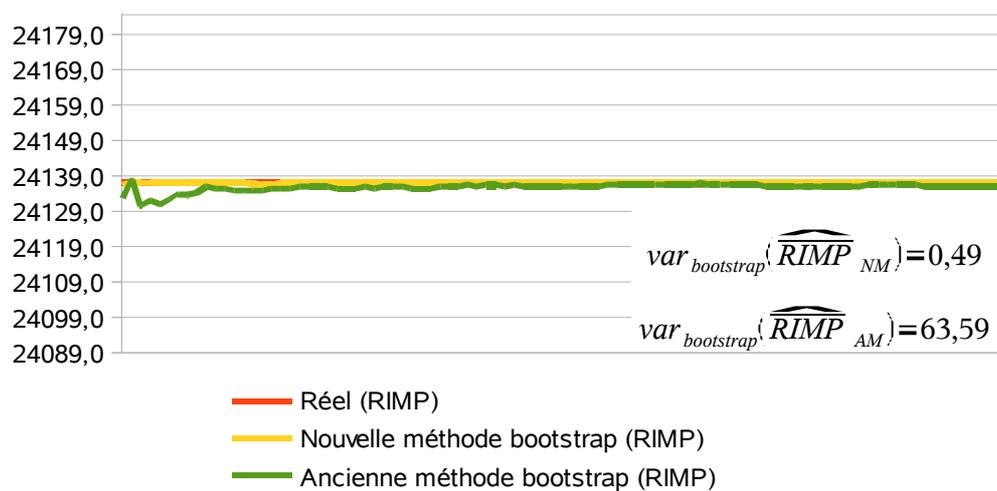
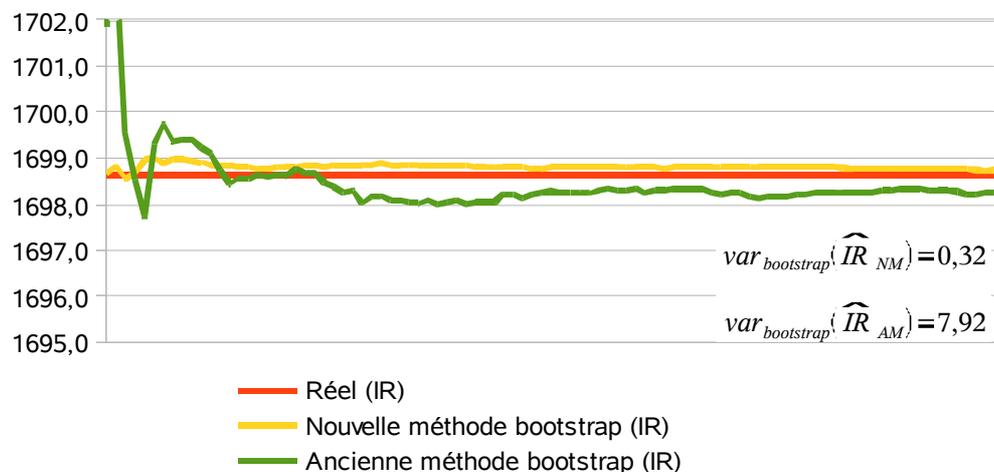
Afin de tester la fiabilité de la nouvelle procédure d'échantillonnage, nous réalisons ici une centaine de tirage bootstrap (95 exactement). En comparant les moyennes réelles de nos principales variables d'intérêt avec leur équivalent bootstrap résultant de l'ancienne et de la nouvelle méthodologie, il est possible d'apprécier graphiquement le gain de précision obtenu.

L'ancienne méthode d'échantillonnage n'étant pas aléatoire (racine fixée pour le premier tirage de chaque strate) elle a, pour les besoins de l'exercice, été modifiée en initialisant aléatoirement le sondage. Cela n'empêche en rien la comparaison, l'échantillon fixiste traditionnel n'étant finalement qu'un cas particulier de l'ensemble des possibilités aléatoires.

Les trois graphiques ci-dessous présentent donc ces trois moyennes pour l'impôt sur le revenu (IR), le revenu fiscal de référence (RFR) et le revenu imposable (RIMP). A noter que l'échelle des graphiques est à chaque fois bornée à +/-0,2 % de la moyenne réelle : cela illustre le fait que l'ancienne méthode comme la nouvelle sont toutes les deux très précises.

¹⁰ L'ordre de grandeur de l'impôt sur le revenu net pour les dernières années est de 70 milliards d'euros.

Graphique 1 - Moyennes réelles et bootstrap de l'impôt, du revenu imposable et du RFR (en euros ; moyennes bootstrap de 1 à 95 échantillons)



Source : POTE 2012-5, LOU 2012-5, FELIN 2012-5

Champ : POTE complet ; 95 versions aléatoires de chaque échantillon.

Note de lecture : l'échelle de l'axe des ordonnées est bornées à +/-0,2% de la moyenne réelle de chaque variable. L'axe des abscisses débute avec un échantillon et se termine avec 95 échantillons cumulés. Les résultats intégrés dans les graphiques indiquent la variance bootstrap des moyennes estimées.

Il ressort nettement de ces trois graphiques que les échantillons tirés avec la macro FastCube sont bien plus précis que l'ancien échantillon. Ce constat est vrai pour les trois variables testées ici. Concernant la variable de revenu imposable, la moyenne bootstrap est même confondue avec la moyenne réelle.

La comparaison des variances estimées par bootstrap est également très satisfaisante pour la nouvelle méthodologie (« NM » sur le graphique). Ainsi, la moyenne estimée de l'impôt a une variance qui passe de 7,92 à 0,32 avec l'amélioration de la méthodologie ; soit une division par 25 du niveau de variance. L'écart est également visible pour le RFR moyen, puisque la variance estimée par bootstrap est divisée par 6, passant de 1124,85 à 176,85. Enfin, le revenu imposable (RIMP) est la variable pour laquelle l'amélioration est la plus importante, mais surtout, où la variance de l'estimateur est la plus faible : 0,49 avec la nouvelle méthode.

Ces résultats empiriques viennent donc valider *a posteriori* les tests et les choix faits au cours de ce travail (cf. 5.2 notamment). Comme attendu, les variances estimées sont bien moindres que celles résultant du calcul théorique avec la formule de Neyman. Ces gains relatifs sont évidemment à associer aux bons niveaux de précision initiaux.

Conclusion

La réflexion sur la méthode d'élaboration des échantillons fiscaux est un sujet permanent de travail au sein du bureau GF-3C. A ce stade et concernant l'impôt sur le revenu, une évolution importante a été réalisée. Le passage à un échantillon équilibré permet désormais de s'assurer, plus encore qu'avant, de la fiabilité des résultats produits au niveau des strates et au niveau macroéconomique. Au regard des travaux d'évaluation menés par le bureau, cette assurance constitue un gain non négligeable. Cette refonte de la méthodologie d'échantillonnage a également été l'occasion de repenser les problématiques de l'impôt sur le revenu. Les paramètres déterminant l'impôt et ses évolutions législatives sont ainsi mieux pris en compte lors de l'échantillonnage.

Cette étude ne vient néanmoins pas clore toutes les réflexions en matière d'échantillonnage. Au contraire, elle ouvre la voie à de nouvelles possibilités. Ainsi, afin de pouvoir traiter de manière aussi fiable les demandes infra-nationales que les demandes nationales, il serait intéressant de réfléchir à la mise en œuvre d'un tirage spatialement équilibré ([10]). Un enjeu serait également de transformer cet échantillon en panel, avec pour objectif de pouvoir suivre et évaluer l'impact des évolutions législatives au niveau microéconomique sur plusieurs années.

Enfin, les réflexions méthodologiques menées pour traiter l'impôt sur le revenu sont également applicables à d'autres impôts. Ainsi, pour les données de taxe d'habitation et de taxe foncière, aucun échantillon n'existait auparavant pour réaliser des simulations. Grâce au travail mené en parallèle avec F. Robin stagiaire Ensai durant l'été 2014, un échantillonnage similaire à celui réalisé sur l'impôt sur le revenu a pu être initié pour ces deux impôts locaux.

Annexes

Annexe 1 - Ecart relatif entre les totaux du POTE et des différents échantillons testés, selon le statut d'imposition

Variables d'équilibrage	Échantillons (foyers imposés)						POTE (total)
	AM	1	2	3	4	5	
Impôt sur le revenu	0,07%	0,02%	0,02%	0,02%	0,06%	0,02%	65 286 000
Revenu fiscal de référence	0,29%	0,05%	0,03%	0,06%	0,04%	0,04%	730 830 000
Revenu imposable	0,21%	0,00%	0,00%	0,00%	0,00%	0,00%	687 510 000
Traitements et salaires	0,36%	0,11%	0,08%	0,07%	0,01%	0,01%	441 960 000
Pensions et rentes	0,30%	0,30%	0,24%	0,22%	0,01%	0,01%	175 270 000
Bénéfices industriels et commerciaux	0,17%	1,20%	0,31%	0,25%	0,32%	0,32%	13 489 000
Bénéfices non commerciaux	0,22%	0,40%	0,08%	0,06%	0,07%	0,08%	31 273 000
Bénéfices agricoles	1,20%	1,72%	0,24%	0,29%	0,25%	0,25%	6 737 363
Revenus fonciers	1,25%	0,31%	0,42%	0,48%	0,05%	0,05%	29 371 000
Revenus de capitaux mobiliers	0,18%	0,63%	0,20%	0,17%	0,22%	0,18%	13 070 000
Divers	0,47%	0,36%	0,36%	0,17%	0,10%	0,14%	10 163 000
Crédits d'impôt	0,20%	0,02%	0,03%	0,02%	0,02%	0,04%	5 023 972
Réductions d'impôt	0,78%	0,11%	0,10%	0,10%	0,09%	0,12%	4 697 639
Charges	0,86%	0,14%	0,20%	0,17%	0,14%	0,23%	8 546 847
Nombre	-0,22 %	0,00%	0,00%	0,00%	0,00%	0,00%	19 052 595

Variables d'équilibrage	Échantillons (foyers non imposés)						POTE (total)
	AM	1	2	3	4	5	
Impôt sur le revenu	0,14%	0,13%	0,14%	0,12%	0,18%	0,19%	-3 200 405
Revenu fiscal de référence	0,66%	0,06%	0,06%	0,07%	0,02%	0,05%	203 640 000
Revenu imposable	0,57%	0,01%	0,01%	0,01%	0,01%	0,01%	194 700 000
Traitements et salaires	0,91%	0,18%	0,13%	0,21%	0,04%	0,03%	129 410 000
Pensions et rentes	0,10%	0,44%	0,21%	0,43%	0,04%	0,03%	71 849 000
Bénéfices industriels et commerciaux	0,98%	2,15%	0,79%	0,87%	1,18%	0,94%	4 307 844
Bénéfices non commerciaux	1,05%	2,19%	1,64%	1,49%	2,36%	1,24%	2 273 561
Bénéfices agricoles	0,92%	4,89%	1,69%	0,98%	1,04%	0,82%	1 666 308
Revenus fonciers	1,20%	2,10%	2,32%	1,92%	0,56%	0,39%	3 631 844
Revenus de capitaux mobiliers	1,00%	2,10%	0,92%	0,78%	0,47%	0,87%	2 179 359
Divers	3,70%	4,49%	7,71%	6,07%	4,75%	4,15%	549 682
Crédits d'impôt	0,79%	0,29%	0,31%	0,17%	0,27%	0,27%	4 741 668
Réductions d'impôt	0,44%	0,15%	0,23%	0,12%	0,11%	0,18%	1 399 313
Charges	3,46%	3,14%	2,21%	3,06%	2,13%	3,70%	2 398 554
Nombre	0,23%	0,00%	0,00%	0,00%	0,00%	0,00%	17 497 315

Source : POTE 2012-5.

Champ : échantillon ancienne méthode (AM), 5 échantillons nouvelle méthode (NM) et POTE complet.

Note de lecture : pour les seuls foyers imposés, l'écart moyen (sur 10 jeux de données) entre le total d'IR de l'échantillon n°1 et le total du POTE est de 0,02%.

Annexe 2 – Validation du nouvel échantillon par comparaison aux données nationales de l'Etat 1507 M à la cinquième émission des revenus 2012 (version diffusable)

Extrait du CADRE I de l'Etat 1507 M pour les revenus 2012 (cinquième émission)			
Bases d'imposition	Effectif	Montant	Écart relatif de l'échantillon
SALAIRES ET ASSIMILES (3)	> 1 000 000	Env. 500 x 10e9	0%
PENSIONS ET RENTES	> 1 000 000	Env. 100 x 10e9	0%
DONT PENSIONS ALIMENTAIRES PERCUES (*)	> 1 000 000	Env. 1 x 10e9	-2%
DEFICITS SALAIRES ET ASSIMILES, PENSIONS ET RENTES	< 10 000	Env. 100 x 10e6	-69%
GAINS DE LEVEE D'OPTION AU QUOTIENT	< 1 000	Env. 10 x 10e6	20%
RENTES VIAGERES A TITRE ONEREUX	< 500 000	Env. 500 x 10e6	-1%
BENEFICES AGRICOLES AU FORFAIT	< 500 000	Env. 500 x 10e6	-2%
BENEFICES AGRICOLES REEL (3)	< 500 000	Env. 10 x 10e9	0%
B.A. DEFICITS IMPUTABLES	< 100 000	Env. 500 x 10e6	-7%
B.I.C. PROF : REGIME MICRO, VENTES	< 500 000	Env. 500 x 10e6	-9%
B.I.C. PROF : REGIME MICRO, SERVICES ET LOCATIONS MEUBLEES	< 500 000	Env. 1 x 10e9	9%
B.I.C. PROF : BENEFICE (3)	< 1 000 000	Env. 10 x 10e9	0%
B.I.C. PROF : DEFICITS	< 100 000	Env. 1 x 10e9	11%
LOCATIONS MEUBLEES PROF : BENEFICE (3)	< 10 000	Env. 10 x 10e6	29%
LOCATIONS MEUBLEES PROF : DEFICITS	< 10 000	Env. 100 x 10e6	6%
B.I.C. NON PROF : REGIME MICRO, VENTES	< 50 000	Env. 10 x 10e6	-23%
B.I.C. NON PROF : REGIME MICRO, SERVICES	< 50 000	Env. 100 x 10e6	6%
B.I.C. NON PROF : BENEFICE (3)	< 50 000	Env. 100 x 10e6	-6%
B.I.C. NON PROF : DEFICITS DE L'ANNEE	< 50 000	Env. 100 x 10e6	38%
B.I.C. NON PROF : DEFICITS ANTERIEURS IMPUTES	< 10 000	Env. 10 x 10e6	6%
LOCATIONS MEUBLEES NON PROF : REGIME MICRO, LOCATIONS MEUBLEES	< 500 000	Env. 1 x 10e9	4%
LOCATIONS MEUBLEES NON PROF : REGIME MICRO, MEUBLES DE TOURISME	< 100 000	Env. 100 x 10e6	-3%
LOCATIONS MEUBLEES NON PROF : BENEFICE (3)	< 100 000	Env. 100 x 10e6	3%
LOCATIONS MEUBLEES NON PROF : DEFICITS DE L'ANNEE	< 100 000	Env. 500 x 10e6	0%
LOCATIONS MEUBLEES NON PROF : DEFICITS ANTERIEURS IMPUTES	< 10 000	Env. 10 x 10e6	-3%
B.N.C. PROF : REGIME SPECIAL	< 500 000	Env. 1 x 10e9	1%
B.N.C. PROF : DECLARATION CONTROLEE (3)	< 1 000 000	Env. 50 x 10e9	0%
B.N.C. PROF : DEFICITS	< 50 000	Env. 500 x 10e6	-18%
B.N.C. NON PROF : REGIME SPECIAL	< 100 000	Env. 100 x 10e6	-12%
B.N.C. NON PROF : DECLARATION CONTROLEE (3)	< 50 000	Env. 500 x 10e6	-4%
B.N.C. NON PROF : DEFICITS DE L'ANNEE	< 10 000	Env. 10 x 10e6	-20%
B.N.C. NON PROF : DEFICITS ANTERIEURS IMPUTES	< 1 000	Env. 10 x 10e6	-24%
R.C.M. : REVENUS (3)	> 1 000 000	Env. 10 x 10e9	0%
REVENUS FONCIERS : REVENUS (3)	> 1 000 000	Env. 10 x 10e9	0%
REVENUS FONCIERS : DEFICITS	< 1 000 000	Env. 1 x 10e9	-1%
MICRO FONCIER	> 1 000 000	Env. 1 x 10e9	0%
P.V. A COURT TERME ACTIVITES NON SALARIEES	< 1 000	Env. 10 x 10e6	49%
M.V. A COURT TERME ACTIVITES NON SALARIEES	< 1 000	Env. 1 x 10e9	48%
SOMMES A AJOUTER AU REVENU	< 1 000	Env. 1 x 10e9	-47%
TOTAUX			
DEFICITS GLOBAUX REPORTABLES	< 100 000	Env. 1 x 10e9	24%
DEFICITS AGRICOLES REPORTABLES	< 50 000	Env. 500 x 10e6	-21%
R.C.M. : DEFICITS REPORTABLES	< 500 000	Env. 500 x 10e6	14%
DEFICITS FONCIERS REPORTABLES	< 1 000 000	Env. 10 x 10e9	-2%
M.V. COURT TERME ACTIVITES NON SALARIEES REPORTABLES	< 100	Env. 100 x 10e6	-100%
D.I.C. NON PROFESSIONNELS REPORTABLES	< 50 000	Env. 500 x 10e6	29%
DEFICITS DES LOCATIONS MEUBLEES NON PROF. REPORTABLES	< 100 000	Env. 1 x 10e9	-3%
DEF. NON COMMERCIAUX NON PROF. REPORTABLES	< 10 000	Env. 100 x 10e6	-37%

Bibliographie

- [1] Ardilly P., Les techniques de sondage, TECHNIP, 2006.
- [2] Bankier M. D., « Power Allocations : Determining sample Sizes for Sub-national Areas », The American Statistician, vol.42, 1988.
- [3] Chauvet G., Tillé Y., « Fast SAS Macros for balancing Samples user's guide », documentation de l'utilisateur, 2005.
- [4] Chauvet, G., et Tillé, Y. « A fast algorithm for balanced sampling », Comp. Statistics, 21, 53-62, 2006.
- [5] Deville, J.-C., Grosbras, J.-M., and Roth, N., « Efficient sampling algorithms and balanced sample », In Verlag, P., editor, COMPSTAT, Proceeding in Computational Statistics, pages 255–266, 1988.
- [6] Deville, J.-C. and Tillé, Y., « Efficient balanced sampling: the cube method. Biometrika », 91:893–912, 2004.
- [7] DGFIP, « Les impôts des particuliers : l'impôt sur le revenu », Annuaire statistique http://www2.impots.gouv.fr/documentation/statistiques/annuaire2012/pages_web/statistiques2012.htm
- [8] DGFIP, « L'impôt sur le revenu par region, département, commune », IRCOM en ligne <http://www2.impots.gouv.fr/documentation/statistiques/ircom2013/ir2013.htm>
- [9] DGFIP, « Présentation du fichier permanent POTE », Documentation technique interne, 2014.
- [10] Grafström, A. Tillé, Y., « Doubly balanced spatial sampling with spreading and restitution of auxiliary totals », Environmetrics, 24: 120–131, 2013.
- [11] Harnois J., « Échantillonnage et mise en œuvre logicielle de calculs de précision et de redressements sur les contribuables à l'impôt sur le revenu », Rapport de stage de deuxième année de scolarité Ensaé, 1997.
- [12] Insee, France, portrait social, 2009.
- [13] Koubi M., Mathern S., « La nouvelle méthode d'échantillonnage de l'enquête trimestrielle ACEMO depuis 2006 », Document d'études, n°146, 2009.
- [14] Rousseau S., Tardieu F., « La macro SAS CUBE d'échantillonnage équilibré », documentation de l'utilisateur. Technical report, INSEE, 2004.