

LA QUALITÉ DES DONNÉES SUR LES REVENUS : ENQUÊTE VERSUS FICHER ADMINISTRATIF

L'EXEMPLE DE L'ENQUÊTE SRCV 2005

Laurence Dauphin

Insee, Direction des Statistiques Démographiques et Sociales

Introduction

SRCV est un panel d'individus constitué à la demande d'Eurostat et enquêté chaque année. Lorsque le dispositif SRCV a été mis en place en 2004, la source de référence nationale pour la mesure des revenus était, et est toujours, l'enquête Revenus Fiscaux (devenue depuis l'enquête revenus fiscaux et sociaux, ERFS), appariement entre les données fiscales et l'enquête Emploi. Dans ERFS, la plupart des revenus sont collectés par appariement avec des données administratives. La France dispose ainsi de deux sources de référence sur les revenus : l'une au niveau national, l'enquête revenus fiscaux et sociaux, l'autre au niveau européen, le dispositif SILC-SRCV. Réconcilier au mieux les deux dispositifs est un enjeu important. Compte tenu des différences de nature entre données administratives et données déclarées dans le domaine des revenus, l'appariement des données SRCV avec les déclarations fiscales est la voie qui paraît la plus naturelle. Ceci doit permettre, de pouvoir utiliser les avantages de chacun des deux types de sources, et ainsi d'améliorer la qualité du dispositif en profitant des atouts des sources administratives.

Celles-ci offrent en effet une qualité homogène et très satisfaisante des données sur les revenus pour la grande majorité de la population. Ce n'est pas le cas pour les sources déclaratives. En effet, les erreurs entre euros et francs, voire même entre euros et anciens francs, et entre montants annuels et mensuels sont courantes, comme le montrent les travaux d'imputation menés sur les données des premières vagues du dispositif SRCV¹. Par ailleurs, certains ménages utilisent des documents, d'autres non. Dans le cas où les documents ne sont pas utilisés, les données sont souvent approximatives ou simplement données en tranches. Enfin, une fraction des ménages ne souhaite répondre à aucune question relative aux revenus. Tous ces éléments conduisent à une hétérogénéité dans la collecte qui nuit aux objectifs de qualité globale, et aux objectifs même du panel qui est de suivre d'une année sur l'autre le revenu d'une même personne (Lollivier et Verger, 2005).

Cependant, les sources administratives ne permettent pas non plus une mesure parfaite des composantes de revenu. Certaines composantes sont exonérées d'impôts et ne sont pas présentes dans les déclarations fiscales. D'autres composantes sont très mal appréhendées, parce que les montants que l'on peut trouver dans les déclarations fiscales résultent de l'application de barèmes fiscaux et ne correspondent pas à une notion de revenu au sens classique du terme.

Un appariement à visée méthodologique entre les données de la collecte SRCV réalisée en mai-juin 2005 et les déclarations fiscales portant sur les revenus 2004 a été effectué.

1. Contexte

1.1 Le dispositif SRCV

Le dispositif SRCV s'inscrit dans le programme d'action communautaire de lutte contre l'exclusion sociale et vise d'abord à obtenir les indicateurs structurels longitudinaux produits pour le rapport de synthèse annuel de la Commission dans les domaines de la répartition des revenus, de la pauvreté et de l'exclusion dans les États membres, en particulier, en ce qui concerne les phénomènes de pauvreté persistante. Il a également pour fin de mettre une base de micro-données sur les revenus et les conditions de vie à disposition des chercheurs, afin qu'ils puissent mener, sur un échantillon

¹ Cf. Rapport intermédiaire sur la qualité EU-SILC 2005.

d'individus suivis dans le temps, des études comparatives et temporelles sur les inégalités et le rôle des politiques sociales et fiscales dans la redistribution.

Le dispositif SRCV donne lieu à une collecte annuelle d'information par voie d'enquête, et le suivi d'individus dans le temps interrogés lors de collectes successives (panel). Chaque année, le panel est alimenté au travers d'un échantillon entrant. Il remplace les ménages issus de l'échantillon « sortant » selon un schéma rotatif. Les ménages du panel sont interrogés pendant 9 ans, les 4 premières années étant obligatoires.

1.2 La collecte des revenus dans SRCV

1.2.1 Structure du questionnaire

L'enquête est composée d'un questionnaire Ménage (destiné à l'ensemble du ménage) et d'un questionnaire individuel posé à toutes les personnes du ménage âgées de 16 ans ou plus (au 1^{er} janvier de l'année d'enquête). Le questionnaire ménage est précédé du Tronc commun des ménages, qui constitue le socle de l'ensemble des enquêtes auprès des ménages conduites par l'Insee.

Un certain nombre de composantes du revenu des ménages est perçu au niveau individuel. Il s'agit :

- des revenus d'activité ou de remplacement (salaires, allocations de chômage ou de préretraite, pensions, retraites ou rentes...),
- de certaines prestations et minima sociaux (prestations maladie, invalidité ou dépendance, bourses d'études, revenu minimum d'insertion, minimum vieillesse, etc.).

En revanche, certaines composantes du revenu des ménages sont difficiles, voire impossibles, à individualiser. Il s'agit :

- des aides au logement et du loyer imputé pour les propriétaires², des prestations familiales, de l'aide sociale ;
- des revenus de la propriété, des transferts entre ménages, des revenus des enfants, des impôts et des remboursements d'intérêts hypothécaires.

Ces composantes sont donc, en général, collectées au niveau du ménage.

Pour limiter les erreurs de collecte, l'équipe de conception a d'une part introduit dans l'outil de collecte CAPI des filtres et des contrôles, d'autre part autorisé une collecte en francs pour de nombreux montants, en particulier lorsque les documents sont anciens (tableau d'amortissement pour la collecte des intérêts d'emprunt, prix de vente du logement).

1.2.2 Exhaustivité et absence de doubles comptes dans la collecte des revenus

Arguant du fait que l'enquêteur n'est pas capable a priori de connaître toutes les sources de revenus du ménage et que ce dernier peut oublier d'en mentionner une, on met en œuvre une stratégie générale de « balayage » des revenus. Cette stratégie consiste à :

- recenser les différents types de revenus perçus par le ménage au cours de l'année de référence (année civile 2004), avant de collecter les montants correspondants,
- vérifier que le montant n'a pas été inclus antérieurement dans un autre revenu pour éviter les double-comptes,
- essayer systématiquement, lorsque l'enquête n'a pas pu ou voulu renseigner un montant, d'obtenir une information en tranches.

1.2.3 Utilisation des documents administratifs et fiscaux

Le questionnement est fonction des documents utilisés. Parmi ces documents, la ou les déclarations des revenus de l'année de référence 2004 (remplies en mars ou avril 2005) sont privilégiées pour les revenus d'activité et de remplacement des adultes, qui sont en quasi-totalité imposables. La déclaration de revenus présente l'avantage d'un mode de recueil des revenus annuels à la fois plus facile et plus fiable. D'une part, les récapitulatifs annuels sont déjà faits. D'autre part, le concept mesuré (revenu déclaré) permet une étude de qualité par comparaison de la distribution des revenus

² Avec le même revenu monétaire un ménage propriétaire de sa résidence principale peut assurer un niveau de vie supérieur à un ménage locataire qui doit, sur ses ressources, payer son loyer. C'est l'objet des loyers imputés que d'estimer cet « avantage » en termes de niveaux de vie du ménage propriétaire par rapport au ménage locataire.

avec celle de l'enquête Revenus fiscaux (enquête de référence à l'Insee pour la distribution des revenus).

Dans le questionnaire individuel, l'enquêté indique combien de déclarations de revenus 2004 il a effectuées, et s'il consent à s'y reporter. Pour les revenus d'activité et de remplacement, le questionnement prend deux formes différentes selon la réponse :

- une boucle (dans CAPI) indexée sur le nombre de déclarations de revenus en cas de consentement,
- sinon, une boucle indexée sur le nombre de sources de revenus (nombre d'employeurs pour les salaires, d'entreprises pour les revenus d'indépendants, de caisses ou organismes de versement pour les retraites ou pensions...).

Dans le premier cas, on obtient un montant annuel de revenu déclaré, qui doit être complété par une interrogation synthétique sur l'existence et le montant de revenus complémentaires non déclarés.

Dans le second cas, on reconstitue un montant annuel de revenu perçu cumulé par source de manière plus analytique : durée de perception, régularité du montant mensuel, et montant mensuel en cas de versements réguliers ou montant annuel en cas de versements irréguliers.

1.3 Les objectifs de la refonte 2008

Le projet de refonte SRCV vise, d'une part, à réorganiser l'ensemble du système d'information sur les ressources et les conditions de vie autour du seul dispositif SRCV, en y incluant des thèmes supplémentaires relatifs aux conditions de vie ; d'autre part, à remplacer, pour une partie des ressources des ménages, la collecte par voie d'enquête par recours aux données administratives, déclarations fiscales et fichiers de prestations des organismes gestionnaires (Cnaf, CCMSA et Cnav). Le recours aux données administratives, données fiscales et sociales, fait donc partie du projet de refonte du dispositif SRCV. Les rapprochements entre données d'enquête et sources administratives, fiscales ou sociales, offrent en effet à plusieurs titres une alternative aux insuffisances d'un mode « tout collecte » auprès des ménages en permettant :

- d'améliorer la qualité du dispositif en profitant des atouts des sources administratives.
 - d'aborder des thèmes qui ne pouvaient l'être jusque là faute de temps.
- En effet, le rapprochement avec des sources administratives doit permettre d'éviter de poser des questions pour lesquelles les données sont déjà disponibles, auprès d'une administration par exemple.
- de répondre à la critique de plus en plus rencontrée, de l'impression désagréable qu'ont certains ménages de donner les mêmes informations à plusieurs administrations.

La refonte du dispositif permettra ainsi d'étudier plus précisément les évolutions des différents indicateurs de pauvreté, et leurs corrélations. Il permettra une analyse annuelle des recouvrements des différents types de pauvreté, pauvreté monétaire, pauvreté en conditions de vie, etc.

1.4 Le test d'appariement

Afin d'évaluer la faisabilité et les risques de la refonte, un test méthodologique a été réalisé à partir des données de l'enquête SRCV 2005.

Le test vise également à expertiser la qualité des données d'enquête et des données administratives obtenues au travers de la déclaration fiscale. Ce test permet de mieux connaître les avantages et inconvénients de ces différentes sources en fonction des variables et populations étudiées. Par ailleurs, il permet de mettre en lumière les difficultés ainsi que les opportunités qu'offrent les appariements.

Ce test d'appariement concerne les personnes interrogées en 2005 dans le cadre de l'enquête SRCV et appartenant à l'échantillon de la collecte 2008. En effet, l'appariement avec les données fiscales de la Direction Générale des Finances Publiques (DGFIP) n'a été annoncé aux ménages qu'à partir de la collecte de mai 2008. L'obligation de collecte loyale nous permet d'apparier les seuls ménages qui en ont été averti.

Cet appariement concerne 5 600 ménages avec un taux de succès de 95 %.

2. Qualité de l'appariement

2.1. Méthodologie

Dans une première phase, l'Insee (pôle «revenus fiscaux » de Rennes) réalise le rapprochement entre déclarations fiscales et fichiers de collecte SRCV selon la technique dite en « double aveugle ». Les fichiers des identifiants des ménages enquêtés SRCV et des foyers fiscaux (FIP) sont rapprochés. Les variables utilisées pour ces rapprochements sont les prénoms, sexe, adresse, date et lieu de naissance. On établit alors une table de passage entre identifiants de collecte SRCV et identifiant FIP, après suppression des autres variables du fichier, en particulier les adresses. Puis on récupère au moyen de cette table les données sur l'impôt, la taxe d'habitation, et les revenus déclarés, et on supprime le numéro FIP.

Dans une dernière phase, l'ensemble des fichiers constitués précédemment sont concaténés au moyen de l'identifiant de collecte SRCV. Puis ce dernier est supprimé du fichier. À l'issue de ces rapprochements, le fichier est un fichier non directement ni indirectement nominatif.

2.2. Taux d'appariement

Lors de ce test, on a cherché à retrouver dans les fichiers fiscaux 14 482 individus (adultes et enfants) dans 5 832 ménages. Au final 13 578 individus ont été effectivement retrouvés dans 5 635 ménages, soit un taux d'appariement de près de 94 % au niveau individuel et de près de 97 % au niveau ménage. A titre de comparaison, le taux d'appariement entre l'enquête Emploi et les fichiers fiscaux pour les revenus 2006 est assez proche : 96 % pour les ménages et 95 % pour les individus.

281 individus ont été retrouvés dans une déclaration d'un membre d'un ménage SRCV (en conjoint ou en personne à charge) mais n'appartiennent pas au ménage SRCV dans l'enquête.

Sur les 13 578 individus appariés, 13 413 ont leurs revenus individualisés complets dans le fichier fiscal. 79 d'entre eux ont connu un changement de situation (mariage, divorce, veuvage) au cours de l'année mais une seule déclaration a été retrouvée au lieu de deux. Les revenus de la 2^e déclaration sont alors imputés en tenant compte de la date de l'événement. 66 individus sont imposés au forfait agricole et leur bénéfice agricole a été estimé par les services de la DGFIP. 20 individus sont imposés au forfait agricole mais leur bénéfice agricole n'a pas été estimé par les services de la DGFIP.

Parmi les 5 635 ménages appariés, 5 382 ont une information pouvant être considérée comme complète et 253 ont une information partielle (il manque les revenus de la déclaration des individus non appariés).

2.3. Les difficultés d'appariement

La principale difficulté rencontrée est la différence entre la notion de ménage utilisée dans les enquêtes de l'INSEE et la notion de foyer fiscal utilisée par l'administration fiscale. En effet, un ménage, est un ensemble de personnes qui partagent le même logement et qui font budget commun. Un couple non marié est donc considéré par l'INSEE comme un seul ménage. Par contre, l'administration fiscale les considérera comme deux foyers fiscaux. Cette difficulté peut être levée en recourant au fichier de la taxe d'habitation qui permet d'observer les couples non mariés, qui font deux déclarations de revenu distinctes, mais paient une seule taxe d'habitation.

Mais certains cas peuvent être plus complexes. Par exemple, certains enfants à charge peuvent figurer sur la déclaration fiscale d'un des parents et habiter avec l'autre. Les étudiants sont le plus souvent rattachés au foyer fiscal de leurs parents tout en habitant un logement indépendant.

Les autres difficultés proviennent de la qualité des données d'identification. L'appariement se fait essentiellement sur l'adresse, le prénom et la date de naissance³. Les prénoms et dates de naissance sont ceux saisis dans CAPI par l'enquêteur lors de la collecte. Les adresses ont fait l'objet d'une saisie informatique a posteriori. Il n'avait pas été précisé en 2005 que ces données pourraient faire l'objet d'un rapprochement avec des fichiers administratifs. Les prénoms saisis ne sont donc pas toujours les mêmes que sur les déclarations fiscales⁴.

³ Le nom crypté sera utilisé à partir de 2008 mais n'a pas pu l'être pour le test.

⁴ En particulier dans le cas de prénoms à consonance étrangère, l'enquêteur peut avoir simplifié son prénom pour aider l'enquêteur.

2.4. Qui sont les individus non appariés ?

904 individus du fichier SRCV 2005 n'ont pas été appariés avec les données fiscales. Pour 206 d'entre eux, il s'agit d'une erreur dans la chaîne de traitement lors du test, qui sera corrigée en 2008. Il s'agit d'enfants à charge qui sont présents à la fois dans le ménage SRCV et dans le foyer fiscal. Ils sont peu nombreux et n'ont pas de revenus ce qui n'interfère donc pas dans les conclusions de ce rapport.

Il reste donc 698 individus (répartis dans 468 ménages) qui n'ont pu être appariés sans qu'il s'agisse d'une erreur de traitement. Pour 361 (52 %) d'entre eux, c'est le ménage (197) dans son ensemble qui n'a pas pu être apparié. Dans ce cas là, il s'agit plus probablement d'un problème sur l'adresse (déménagement en particulier) ou d'une absence de déclaration fiscale⁵.

2.4.1. Non appariement au niveau ménage

Tableau 1 : modélisation logistique du non appariement au niveau ménage

	Probabilité de ne pas être apparié	Rapport des risques (Odd Ratio)
Référence	0,10	
	Écart à la référence en points	
Pas de déménagement récent	-0,05	0,24
Déménagement récent	ref	ref
Habitat collectif	0,06	2,65
habitat indépendant	ref	ref
Naissance de la personne de référence à l'étranger	0,09	4,01
Naissance de la personne de référence en France	ref	ref
Personne seule	ref	ref
Famille monoparentale	ns	ns
Couple sans enfant	-0,05	0,34
Couple avec au moins un enfant	-0,07	0,21
Autre type de ménage	0,31	4,53

Variables explicatives du modèle : déménagement en 2004, type d'habitat, pays de naissance de la personne de référence, type de ménage.
Situation de référence : ménage d'une seule personne ayant déménagé en 2004 dans un logement indépendant, dont la personne de référence est née en France.
Lecture : un ménage qui n'a pas déménagé récemment a 5 % de chances de ne pas être appariés. Quand un ménage a déménagé récemment, son risque de ne pas être apparié est multiplié par 4 (1/0,24).
Tous les coefficients indiqués sont significatifs au seuil de 1 %.
ns = non significatif.
Sortie SAS de la régression logistique en annexe.

Les ménages qui n'ont pas pu être appariés sont plus souvent des ménages qui ont déménagé récemment. Leur adresse dans l'enquête ne correspond alors pas à leur adresse dans les fichiers fiscaux. Les ménages habitants dans des logements indépendants sont plus faciles à retrouver que ceux qui habitent dans des logements collectifs. Plus la taille du ménage est importante, plus on a de chance de retrouver au moins un de ses membres dans les fichiers administratifs.

2.4.2. Non appariement au niveau individu

Certains individus, tout en vivant dans un ménage dans lequel certains membres ont été appariés, n'ont pas pu être retrouvés. Il peut s'agir, par exemple, de couples non mariés pour lesquels, une seule déclaration a pu être retrouvée et dont la deuxième déclaration est faite à une autre adresse. Le non appariement au niveau individuel résulte principalement de l'écart de définition entre le ménage SRCV et le foyer fiscal.

⁵ En France, le taux de couverture fiscale est élevé (de l'ordre de 97%). En 2005, plus de 58 millions de personnes étaient présentes sur une déclaration fiscale, alors que 59 millions de personnes vivent en ménages ordinaires (champ de l'enquête SRCV).

Afin d'isoler ce qui est dû à un effet individu de ce qui est dû à un effet ménage, le modèle suivant porte sur les individus qui vivent dans un ménage apparié⁶ ou qui vivent seul dans leur ménage⁷.

Tableau 2 : modélisation logistique du non appariement au niveau individu

<i>Individus appartenant à un ménage apparié ou vivant seul</i>	Probabilité de ne pas être apparié	Rapport des risques (Odd Ratio)
Référence	0,17	
	Écart à la référence en points	
Déclare ne pas avoir fait de déclaration fiscale	0,10	3,3
Déclare avoir fait au moins une déclaration fiscale	Ref	Ref
Ménage entrant (pas d'enquête l'année précédente)	-0,06	1,4
Fait toujours partie des habitants du ménage (en faisait partie à l'enquête précédente)	Ref	Ref
Fait désormais partie des habitants du ménage (n'en faisait pas partie à l'enquête précédente)	0,29	9,7
Réside aussi ailleurs	0,15	5,4
Ne réside pas ailleurs	Ref	Ref
0 - 15 ans*	ns	ns
15 - 20 ans	-0,07	0,6
20 - 25 ans	0,09	1,9
25 - 30 ans	0,20	3,2
30 - 40 ans*	ns	ns
40 - 45 ans	Ref	Ref
45 - 75 ans*	ns	ns
Plus de 75 ans	ns	ns
Né à l'étranger	-0,03	0,6
Né en France	Ref	Ref
Célibataire	Ref	Ref
Marié	-0,10	0,2
Veuf	0,13	1,4
Divorcé	ns	ns
En couple avec une personne du logement ⁸	-0,07	0,6
En couple avec une personne extérieure au logement	0,13	2,4
Pas en couple	Ref	Ref

Variables explicatives du modèle : avoir rempli une déclaration d'impôts, mouvement individu, autre logement, âge, pays de naissance, situation matrimoniale, vie en couple.

Situation de référence : personne déclarant avoir rempli une déclaration d'impôt, n'ayant pas changé de ménage par rapport à l'année précédente, ne résidant pas aussi ailleurs, ayant entre 40 et 45 ans, née en France, célibataire et ne vivant pas en couple.

Lecture : un individu ayant toutes les caractéristiques de référence et déclarant ne pas avoir rempli de déclaration d'impôt a 27 % de chances de ne pas être appariés, c'est-à-dire 10 points de plus qu'un individu déclarant en avoir rempli une. Quand un individu déclare ne pas avoir rempli de déclaration, son risque de ne pas être apparié est multiplié par 3,3.

**Modalités regroupées dans le tableau mais pas dans la régression.*

Tous les coefficients indiqués sont significatifs au seuil de 1 %.

ns = non significatif.

Sortie SAS de la régression logistique en annexe.

Les individus déclarant ne pas avoir rempli de déclaration fiscale ont 3 fois plus de chance de ne pas être appariés. Cependant, il s'agit d'une déclaration de la part de l'individu qui ne correspond pas toujours à la réalité. En effet, 13 % des individus de 16 ans et plus déclarent dans l'enquête SRCV ne

⁶ Au moins un membre du ménage SRCV a pu être apparié.

⁷ Le ménage n'a alors pas pu être apparié pour des motifs pouvant être individuels.

⁸ Certaines personnes déclarent ne pas vivre en couple, tout en étant mariées.

pas avoir rempli de déclaration fiscale, mais pour 83 % d'entre eux, une déclaration a pu être retrouvée. Il s'agit par exemple, d'enfants de plus de 16 ans qui déclarent ne pas avoir rempli de déclaration en oubliant qu'ils figurent sur celle de leurs parents. Au total, seuls 2 % des individus ont déclaré ne pas avoir rempli de déclaration et n'ont pu être appariés. Ce qui correspond au taux attendu, à la vue des informations fournies par l'administration fiscale (cf. 2.4).

Les individus qui viennent de changer d'adresse ont également moins de chances d'être retrouvés. Il peut s'agir par exemple de mises en couple récentes pour lesquelles l'un des conjoints n'a pas encore communiqué sa nouvelle adresse à l'administration.

Les jeunes entre 16 et 25 ans sont également plus difficiles à retrouver. En effet, les jeunes de plus de 18 ans peuvent déclarer leurs revenus avec leurs parents, sans condition jusqu'à 21 ans et sous conditions jusqu'à 25 ans. Dans le cas de jeunes qui vivent dans un ménage apparié mais qui n'ont pas pu être retrouvés, il peut s'agir d'enfants qui vivent avec l'un de leurs parents mais sont sur la déclaration fiscale de leur autre parent. Le ménage où ils vivent est alors apparié mais pas celui avec lequel ils font leur déclaration. Ce cas est prévu dans l'enquête SRCV 2008 qui fera l'objet du premier appariement. Ces jeunes devront alors donner leurs revenus lors de l'enquête.

Ce problème se pose également pour les enfants qui vivent avec l'un de leurs parents mais figurent sur la déclaration de l'autre parent. Ces enfants ont moins de 16 ans et ne remplissent donc pas de questionnaire individuel. Ils n'ont pas de revenus individuels. Leur non-appariement ne devrait donc pas poser de problème.

Les jeunes qui vivent dans des ménages qui n'ont pas été appariés sont plus certainement des jeunes qui font leur déclaration d'impôts avec leurs parents mais qui habitent dans un logement indépendant. Ils sont alors enquêtés dans leur logement et non celui de leurs parents. Ce cas là est également prévu dans le questionnaire de la collecte 2008.

Les 698 individus non appariés peuvent se répartir en 6 groupes en fonction de leur statut, du statut d'appariement de leur ménage et du nombre de déclarations d'impôts qu'ils déclarent avoir rempli.

Tableau 3 : répartition des individus non appariés

		SO / NR	0	1	2	Total
Salariés	men app	1	20	60	1	82
	men non app	2	32	99	0	133
	Ensemble	3	52	159	1	215
Indépendants	men app	0	3	0	0	3
	men non app	0	14	0	0	14
	Ensemble	0	17	0	0	17
Chômeurs	men app	0	7	9	0	16
	men non app	0	6	13	0	19
	Ensemble	0	13	22	0	35
Retraités	men app	0	1	11	0	12
	men non app	0	2	38	0	40
	Ensemble	0	3	49	0	52
16-25 ans	men app	1	72	21	0	94
	men non app	0	44	11	1	56
	Ensemble	1	116	32	1	150
enfants	men app	130	0	0	0	130
	men non app	99	0	0	0	99
	Ensemble	229	0	0	0	229
Total		233	201	262	2	698

2,5 %

6 %

Lecture : 60 individus non appariés sont salariés, déclarent avoir rempli une déclaration d'impôt et appartiennent à un ménage apparié.

Les enfants de moins de 16 ans ne répondent pas à la question sur le nombre de déclaration et ont donc tous la modalité « SO/NR⁹ »

6 % des individus n'ont pas pu être appariés lors de ce test. Les enfants non appariés n'ont pas de revenus et ne remplissent pas de questionnaire individuel. Leur non appariement est donc sans incidence sur la qualité des données collectées. Le cas des jeunes entre 16 et 25 ans a été prévu dans le questionnaire 2008 qui fait suite à la refonte. Les jeunes qui ont fait leur déclaration à une autre adresse que celle de l'enquête se voient poser toutes les questions concernant leurs revenus.

⁹ Sans Objet / Non Réponse.

Les cas de non appariement qui peuvent avoir une incidence sur la qualité de l'enquête représentent donc seulement 2,5 % des individus. 27 % d'entre eux déclarent ne pas avoir rempli de déclaration et 65 % appartiennent à des ménages non appariés. Ce ne sont donc pas leurs caractéristiques individuelles qui sont à l'origine de cet échec d'appariement.

2.5. Comparaison des revenus des individus appariés et des individus non appariés

Le tableau suivant présente les revenus individuels déclarés, par les individus de 16 ans ou plus, lors de l'enquête SRCV 2005. Respectivement par les individus qui ont pu être appariés et par ceux qui n'ont pas pu l'être, par type de revenu.

Tableau 4 : Comparaison des revenus des individus appariés et des individus non appariés

	Minimum	Q1	Médiane	Q3	Maximum	Moyenne	
Appariés (10 630)	Salaires (PY010N)	0	0	2 145	16 241	225 791	9 453
	Bénéfices indépendants (PY050N)	0	0	0	0	260 000	1 068
	Chômage (PY090N)	0	0	0	0	40 342	492
	Retraites (PY100N)	0	0	0	2 160	280 000	3 752
	Réversion (PY110N)	0	0	0	0	240 000	112
	Maladie (PY120N)	0	0	0	0	54 900	133
	Handicap (PY130N)	0	0	0	0	60 000	209
	Bourses (PY140N)	0	0	0	0	15 000	19
	Somme revenus individuels	0	6 868	13 200	19 927	280 000	15 237
	Niveau de vie (rev dispo par uc)	-209 101	11 549	15 885	21 655	188 899	18 117
	Age	17	33	46	62	101	48
Non-appariés (305)	Salaires (PY010N)	0	0	0	3 935	67 713	366
	Bénéfices indépendants (PY050N)	0	0	0	0	10 000	43
	Chômage (PY090N)	0	0	0	0	60 000	1 198
	Retraites (PY100N)	0	0	0	0	16 481	376
	Réversion (PY110N)	0	0	0	0	158 400	396
	Maladie (PY120N)	0	0	0	0	6 648	22
	Handicap (PY130N)	0	0	0	0	8 544	161
	Bourses (PY140N)	0	0	0	0	3 900	80
	Somme revenus individuels	0	0	900	8 054	158 400	5 941
	Niveau de vie (rev dispo par uc)	4 242	10 221	12 892	18 267	121 807	16 128
	Age	17	20	23	32	96	30

Source : SRCV 2005

Les individus qui n'ont pu être appariés sont beaucoup plus jeunes que la moyenne. Ils ont en général moins de 30 ans. Leurs revenus sont très faibles. 30 % d'entre eux ont déclaré des salaires lors de l'enquête SRCV 2005. 1 % ont déclaré des revenus d'indépendants, 10 % des allocations chômage, 5 % des retraites et 5 % des bourses étudiantes. 42 % d'entre eux n'ont aucun revenu individuel (contre 11 % des individus appariés).

Les individus non appariés gagnent en moyenne 6 000 euros par an, alors que les individus qui ont pu être appariés déclarent gagner 15 000 euros, par an, en moyenne. Les revenus déclarés par les individus de 16 ans et plus non appariés représentent 1,0 % des revenus individuels déclarés dans l'enquête (2,5 % des individus).

3. Impact sur les données transversales

3.1. Impact sur le niveau de vie

3.1.1. Impact sur la distribution

On compare ici le niveau de vie des ménages pour lesquels l'information est complète, calculé avec les données fiscales et celui calculé avec les données d'enquête. L'échantillon utilisé étant un sous-échantillon¹⁰ de celui de l'enquête SRCV 2005, un recalage des données a été effectué avec CALMAR.

L'écart entre le niveau de vie médian calculé avec les données d'enquête avant redressement et celui calculé avec les données DGFIP est de l'ordre de -5 %. Le niveau de vie médian calculé avec les données d'enquête est inférieur à celui calculé avec les données fiscales. Après imputation, l'écart relatif est moins important, de l'ordre de -2 %. L'importance de l'écart constaté entre données brutes et données appariées à la source fiscale est à relativiser en raison de la nature des biais de déclaration qui peuvent survenir au moment de la collecte de l'enquête. En effet, ces biais portent sur des confusions qui peuvent ne pas être liées à une sous-déclaration des revenus, mais plutôt à des confusions mensuel / annuel ou euros / francs.

Tableau 5 : Distribution du revenu disponible par unité de consommation

Centiles	5 328 ménages			
	Enquête avant imputation	Enquête après imputation	DGFIP*	DGFIP**
Min	-253 009	-209 101	-206 364	-38 680
1 %	-2 034	3 217	2 119	2 216
5 %	3 489	7 113	6 722	6 895
10 %	6 864	8 549	8 622	8 685
20 %	9 360	10 400	10 769	10 823
30 %	11 209	12 084	12 648	12 694
40 %	12 990	13 677	14 053	14 137
50 %	14 953	15 549	15 753	15 861
60 %	17 019	17 628	17 679	17 780
70 %	19 603	19 865	19 878	19 893
80 %	22 822	23 063	23 071	23 242
90 %	29 502	29 281	28 759	28 739
95 %	36 286	35 782	35 129	35 583
99 %	59 817	55 393	54 437	54 328
Max	188 899	188 899	582 499	307 160
Moyenne	16 954	17 733	17 750	17 839
Taux de pauvreté	18,3 %	13,0 %	13,3 %	13,2 %

*Salaires, retraites et allocations chômage sont ceux déclarés auprès de l'administration fiscale.

Les impôts, les bénéfices des indépendants, les revenus fonciers, les minima sociaux et les allocations sont ceux de l'enquête.

**En intégrant les impôts récupérés auprès de l'administration fiscale

L'impact de l'appariement sur la forme de la distribution du revenu disponible¹¹ est faible si on considère les données obtenues après imputation. En effet, les imputations sont réalisées de façon à obtenir des résultats cohérents avec l'enquête ERF¹², résultant d'un appariement de l'enquête Emploi et des données fiscales. Les écarts les plus importants se situent dans les queues de distribution. La forme de la distribution du niveau de vie calculé à partir des deux sources, confirme une tendance à la

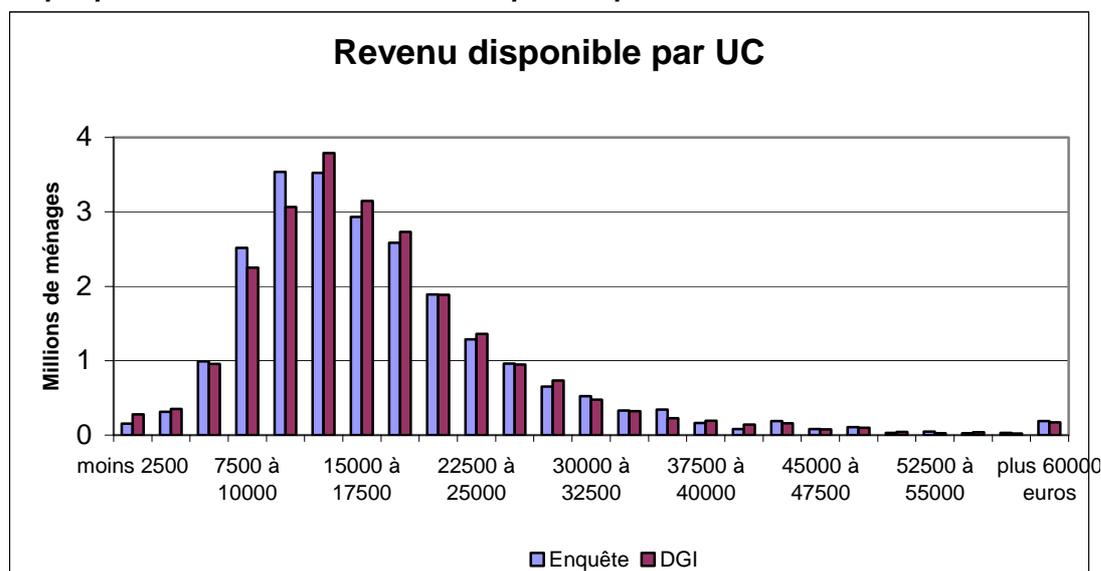
¹⁰ Enquêtés SRCV 2005 réinterrogés en 2008.

¹¹ Revenu disponible = somme des revenus perçus - impôts versés. Le revenu disponible peut être négatif dans le cas d'un revenu faible et d'un impôt élevé ou dans le cas de revenus négatifs (bénéfice des indépendants, revenus fonciers ou financiers).

¹² Enquête Revenus Fiscaux.

sous-évaluation des revenus dans l'enquête. En effet, les ménages ont tendance à oublier une partie de leur rémunération (primes en particulier) ou à en arrondir les montants. Cette sous-déclaration était en partie corrigée lors des travaux d'imputation en aval de la collecte.

Graphique 1 : Distribution du revenu disponible par unité de consommation



3.1.2. Impact sur les taux de pauvreté après transferts sociaux

Tableau 6 : Taux de pauvreté en fonction de la source utilisée

	13 447 individus ¹³			
	Enquête avant imputation	Enquête après imputation	DGFIP*	DGFIP**
Taux de pauvreté	18,1 %	13,0 %	12,5 %	13,4 %
Seuil de pauvreté ¹⁴	9 037	9 434	9 513	9 551

*Salaires, retraites et allocations chômage sont ceux déclarés auprès de l'administration fiscale.

Les impôts, les bénéfices des indépendants, les revenus fonciers, les minima sociaux et les allocations sont ceux de l'enquête.

**En intégrant les impôts récupérés auprès de l'administration fiscale.

Les écarts constatés sur les niveaux de vie entraînent une diminution du taux de pauvreté¹⁵, très nette lorsqu'il est comparé aux données d'enquête non redressées (-4,7 points), faible (+0,4 point) lorsque la comparaison porte sur les données redressées. Ces résultats sont à mettre en perspective avec l'intervalle de confiance autour du seuil de pauvreté dans SRCV, de plus ou moins 0,5 point¹⁶. L'écart constaté ici est donc inférieur à la précision habituellement calculée : l'impact de l'appariement n'est donc pas de grande ampleur.

Le taux de pauvreté après transferts sociaux est défini comme la proportion de personnes ayant un niveau de vie inférieur au seuil de pauvreté. Tous les individus d'un ménage ont le même niveau de vie, qui est égal au revenu disponible divisé par le nombre d'unités de consommation du ménage. Le seuil de pauvreté est fixé à 60 % de la médiane du niveau de vie des individus.

¹³ Individus, y compris enfants, appartenant à un ménage dans lequel tous les individus ont pu être retrouvés.

¹⁴ Dans l'enquête SRCV 2005, le seuil de pauvreté était de 9 562 euros par an, soit un taux de pauvreté de 13%.

¹⁵ Le taux de pauvreté est le pourcentage de la population dont le niveau de vie est inférieur au seuil de pauvreté. Le seuil retenu ici est égal à 60 % du niveau de vie médian de l'ensemble des individus.

¹⁶ Cf. rapport intermédiaire sur la qualité SRCV 2005.

3.1.2.1. Taux de pauvreté selon l'âge et le sexe

Tableau 7 : Taux de pauvreté selon l'âge et le sexe

Taux de pauvreté	Total			0 - 15			0 - 64			16+			16 - 64		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Données administratives	13,4	13,9	12,8	16,3	17,7	14,7	13,7	14,2	13,1	12,5	12,7	12,3	12,7	12,8	12,5
Données d'enquête	13,0	13,8	12,0	14,0	15,0	12,7	12,3	12,8	11,8	12,7	13,4	11,8	11,8	12,0	11,5

Taux de pauvreté	16 - 24			25 - 49			50 - 64			65+		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Données administratives	18,4	17,9	18,9	12,4	13,0	11,8	10,1	10,4	9,9	11,7	12,3	11,0
Données d'enquête	18,3	17,5	19,1	10,4	11,0	9,8	10,9	11,2	10,6	16,1	18,2	13,3

L'écart entre le taux de pauvreté calculé avec les données d'enquête corrigées et celui calculé avec les données fiscales est inférieur à la largeur de l'intervalle de confiance. Les écarts les plus importants s'observent pour les moins de 16 ans et pour les plus de 65 ans.

En effet, les retraités ont tendance à omettre une partie de leurs revenus lors de l'enquête. La rectification opérée avec l'appariement conduit à un taux de pauvreté plus bas pour les retraités lorsque l'on calcule le revenu disponible du ménage en intégrant les revenus déclarés auprès de l'administration fiscale.

3.1.2.2. Taux de pauvreté selon le type de ménage

Tableau 8 : Taux de pauvreté selon le type de ménage

Taux de pauvreté	Sans enfants : total	Personnes seules (total)	2 adultes, moins de 65	2 adultes, dont au moins un de 65 ou +	Avec enfants : total	Familles monoparentales	2 adultes, 1 enfant et +
Données administratives	11,0	16,2	8,5	10,2	15,6	20,9	14,3
Données d'enquête	12,8	18,7	8,5	12,6	13,7	22,6	11,6

L'écart important, entre le taux de pauvreté calculé avec les données d'enquête et celui calculé avec les données fiscales, observé pour les moins de 16 ans se retrouve pour les ménages avec enfants. Le taux de pauvreté des couples avec enfants est plus élevé lorsque l'on utilise les données fiscales. Pour les familles monoparentales, on observe le phénomène inverse. Ces écarts importants sur les taux de pauvreté correspondent cependant plus à une instabilité de cet indicateur qu'à un réel écart dans les revenus. En effet, le seuil de pauvreté correspond à une concentration importante de ménages et un écart de quelques euros peut modifier de plusieurs points le taux de pauvreté.

Tableau 9 : Niveau de vie selon le type de ménage

Niveau de vie	Personnes seules (total)	2 adultes, moins de 65	2 adultes, dont au moins un de 65 ou +	Familles monoparentales	2 adultes, 1 enfant et +
Données administratives	16 424	21 327	17 942	14 298	17 877
Données d'enquête	15 894	21 562	17 931	13 871	18 093
Effectif non pondéré	1488	870	729	276	1557

Les écarts de niveau de vie observés pour les personnes seules et pour les familles monoparentales sont de l'ordre de 3 %. Pour les autres types de ménages, ces écarts sont inférieurs ou égaux à 1 %. Le taux de pauvreté des personnes seules est moins élevé lorsqu'il est calculé avec les données fiscales. On retrouve, ici, le phénomène observé pour les plus de 65 ans et les retraités.

¹⁷ Il est également possible que certains enquêteurs, ayant affaire à des retraités, ne posent pas les questions sur les salaires.

3.1.2.3. Taux de pauvreté selon le statut professionnel le plus fréquent et le sexe

La décomposition du taux de pauvreté selon le statut professionnel confirme le constat déjà évoqué d'une sous-estimation par enquête du taux de pauvreté des retraités.

Tableau 10 : Taux de pauvreté selon le statut professionnel

Taux de pauvreté	Total			En emploi			Sans emploi : total			Sans emploi : chômeurs		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Données administratives	12,5	12,7	12,3	7,8	7,1	8,5	17,6	17,6	17,5	23,7	20,0	28,3
Données d'enquête	12,7	13,4	11,8	6,1	5,4	6,8	19,8	20,5	18,7	24,5	20,7	29,3

Taux de pauvreté	Sans emploi : retraités			Sans emploi : autres inactifs		
	Total	F	H	Total	F	H
Données administratives	9,7	9,5	10,0	33,3	29,9	49,6
Données d'enquête	13,6	13,9	13,2	31,7	31,6	31,7

Sont exclus du champ les personnes de moins de 16 ans au 1^{er} janvier de l'année de l'enquête et les individus qui n'ont pas gardé le même statut d'occupation pendant plus de la moitié de l'année.

3.1.3. Impact individuel

3.1.3.1. Passage du seuil de pauvreté

Le taux de pauvreté ainsi calculé est de 13,4 % avec les données administratives et de 13,0 % avec les données d'enquête. Les deux taux sont donc très proches.

Tableau 11 : Statut face à la pauvreté en fonction de la source

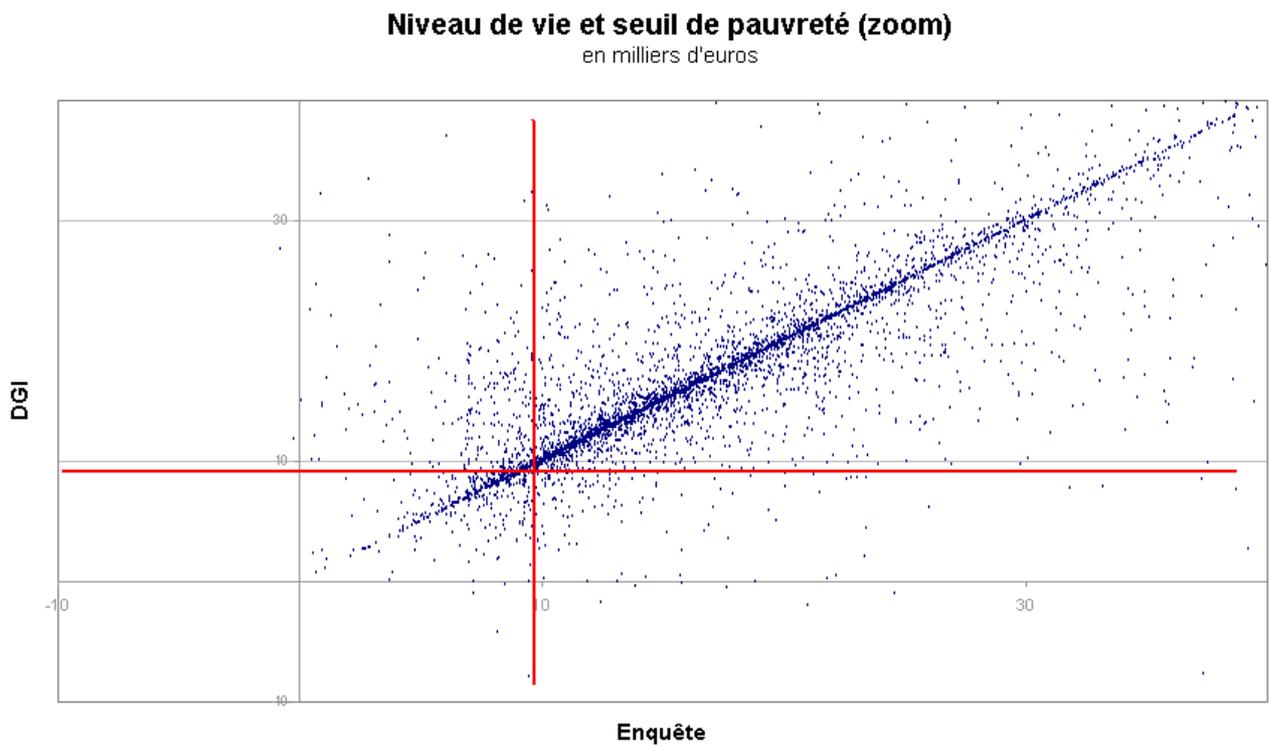
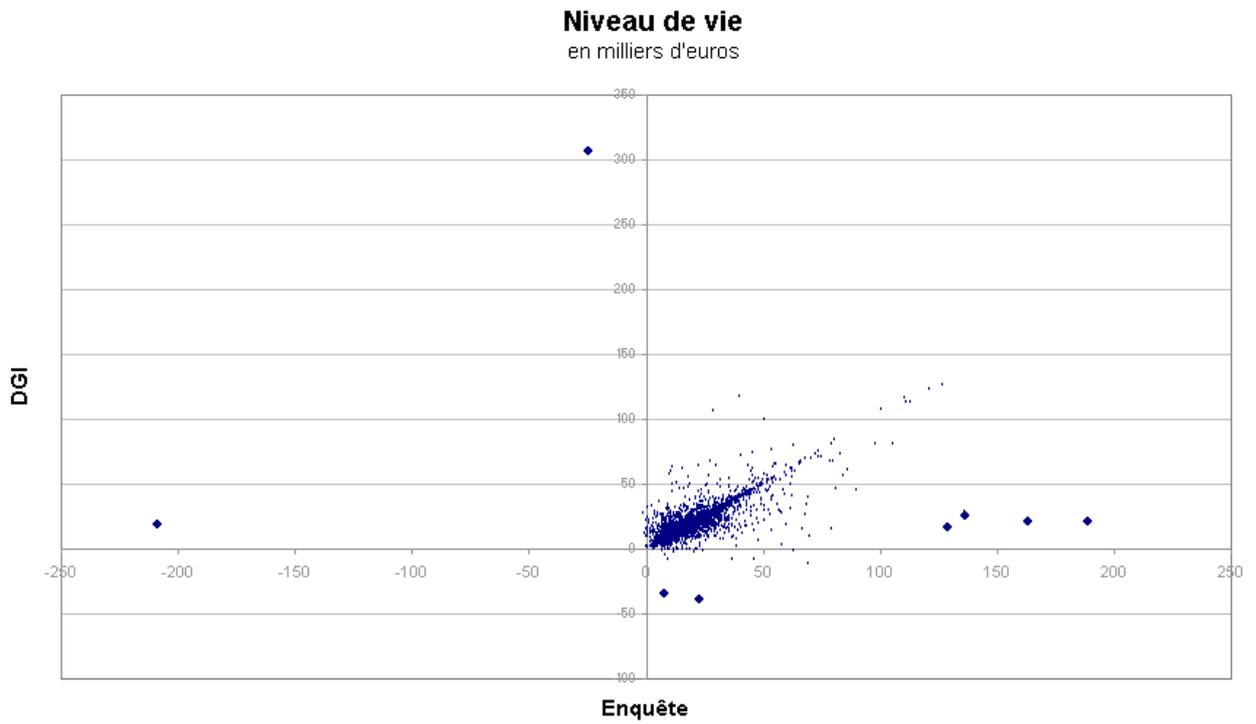
Nombre d'individus (non pondéré)		Données d'enquête		
		Non pauvre	Pauvre	Total
Données fiscales	Non pauvre	11 246	497	11 743
	Pauvre	540	1 164	1 704
	Total	11 786	1 661	13 447

Pourcentage (pondéré)		Données d'enquête		
		Non pauvre	Pauvre	Total
Données fiscales	Non pauvre	82,7 %	3,9 %	86,6 %
	Pauvre	4,3 %	9,0 %	13,4 %
	Total	87,0 %	13,0 %	100 %

Le nombre de personne vivant en dessous du seuil de pauvreté est quasiment le même avec les deux sources. Cependant certaines personnes étaient en dessous du seuil de pauvreté avec les données d'enquête et ne le sont plus avec les données fiscales, et vice-versa. Environ 4 % des individus passent en dessous du seuil de pauvreté et 5 % passent au-dessus.

Les écarts entre données fiscales et données d'enquête, faibles au niveau macro, sont donc non négligeables au niveau micro. Pour certains ménages, les écarts sont très importants. Il s'agit d'erreurs non corrigées dans l'enquête ou à contrario de valeurs extrêmes corrigées à tort. Ces écarts, bien que de plus centaines des milliers d'euros, sont très exceptionnels et n'ont pas d'incidence sur la qualité globale de l'information recueillie.

Graphiques 2 & 3 : Niveau de vie en fonction de la source de données utilisée



3.1.3.2. Changement de décile de niveau de vie

Tableau 12 : Décile de niveau de vie en fonction de la source

Nombre d'individus (non pondéré)		Données d'enquête										
		1 ^{er} décile	2 ^e décile	3 ^e décile	4 ^e décile	5 ^e décile	6 ^e décile	7 ^e décile	8 ^e décile	9 ^e décile	10 ^e décile	Total
Données fiscales	1 ^{er} décile	798	162	85	46	28	18	10	21	12	25	1 205
	2 ^e décile	187	734	275	49	36	20	12	22	16	13	1 364
	3 ^e décile	46	115	696	321	68	34	32	21	7	6	1 346
	4 ^e décile	42	48	104	594	235	55	25	13	23	3	1 142
	5 ^e décile	31	53	43	141	725	262	23	40	14	40	1 372
	6 ^e décile	58	30	34	51	119	688	208	97	24	30	1 339
	7 ^e décile	36	22	17	22	61	118	769	167	74	38	1 324
	8 ^e décile	16	20	34	14	29	44	142	861	231	41	1 432
	9 ^e décile	29	16	19	19	39	20	56	123	974	155	1 450
	10 ^e décile	8	15	10	6	14	13	29	42	136	1 200	1 473
Total		1 251	1 215	1 317	1 263	1 354	1 272	1 306	1 407	1 511	1 551	13 447

Pourcentage (pondéré)		Données d'enquête										
		1 ^{er} décile	2 ^e décile	3 ^e décile	4 ^e décile	5 ^e décile	6 ^e décile	7 ^e décile	8 ^e décile	9 ^e décile	10 ^e décile	Total
Données fiscales	1 ^{er} décile	6,4	1,2	0,5	0,4	0,2	0,4	0,4	0,1	0,3	0,1	10,0
	2 ^e décile	1,1	6,1	0,8	0,6	0,7	0,2	0,2	0,2	0,2	0,1	10,0
	3 ^e décile	0,8	1,6	5,6	0,9	0,3	0,3	0,2	0,2	0,1	0,0	10,0
	4 ^e décile	0,7	0,4	2,0	5,1	1,0	0,3	0,2	0,0	0,2	0,0	10,0
	5 ^e décile	0,1	0,2	0,4	1,9	5,4	1,0	0,4	0,3	0,3	0,2	10,0
	6 ^e décile	0,1	0,1	0,3	0,6	1,8	5,3	0,9	0,4	0,3	0,2	10,0
	7 ^e décile	0,2	0,1	0,3	0,2	0,2	1,6	5,7	1,2	0,3	0,4	10,0
	8 ^e décile	0,2	0,2	0,1	0,1	0,2	0,7	1,2	6,0	1,0	0,3	10,0
	9 ^e décile	0,2	0,1	0,1	0,2	0,0	0,2	0,5	1,5	6,2	1,1	10,0
	10 ^e décile	0,2	0,1	0,0	0,0	0,2	0,1	0,2	0,3	1,1	7,7	10,0
Total		10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0	100,0	

La diagonale du tableau de contingence qui croise les distributions par décile entre données d'enquête et données fiscales est chargée. 60 % des individus restent dans le même décile de niveau de vie quelles que soient les données utilisées. 82 % ne s'écartent pas de plus d'un décile de leur décile d'origine. 6 % s'écartent de plus de 3 déciles de leur décile d'origine. Pour une grande majorité des individus, le recours aux données administratives est sans incidence sur leur décile de niveau de vie. Pour une minorité, l'impact est non négligeable.

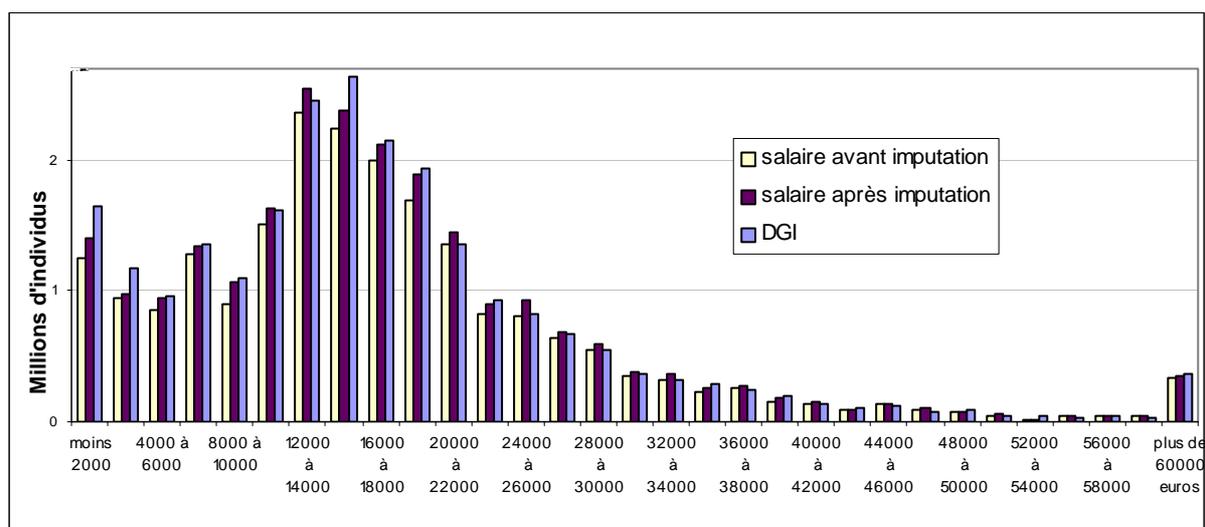
3.2. Impact sur le salaire net (PY010N)

3.2.1. Distribution

L'impact de l'appariement sur la distribution des salaires est assez faible si on considère les valeurs obtenues après les travaux d'imputation¹⁸. L'information sur les salaires est bien collectée par l'enquête et les modèles économétriques, mis en œuvre en aval, lors de la phase d'imputations, fonctionnent bien. Il s'agit d'une information connue du ménage, facile à trouver et facile à modéliser.

¹⁸ Calcul des déciles en annexe.

Graphique 4 : Distribution des salaires en fonction de la source



Le salaire annuel net médian est de 14 898 euros avant imputations, 15 774 euros après imputations et 15 672 euros avec les données provenant des déclarations fiscales. L'écart est donc de 5 % entre les salaires avant imputations et les données fiscales et de 0,6 % entre les données après imputations et les données fiscales. Les grandes caractéristiques de la distribution des salaires sont très proches.

Tableau 13 : Distribution des salaires en fonction de la source

5 396 individus	Minimum	Q1	Médiane	Q3	Maximum	Moyenne
Salaire avant imputation	0	8 725	14 898	21 175	225 791	16 737
Salaire après imputation	28	10 911	15 774	21 890	225 791	18 016
Salaire DGFIP	6	11 162	15 672	21 657	225 791	17 942

Champ : SRCV 2005 - DGFIP, individus qui ont à la fois des salaires après imputation des données d'enquête et dans les fichiers fiscaux.

3.2.2. Impact individuel

L'impact de l'appariement sur la distribution des salaires est faible. Par contre, les écarts au niveau individuel peuvent être importants.

3.2.2.1. Déclarer ou non des salaires

Les écarts les plus importants se trouvent dans le fait de déclarer ou non des salaires. En effet, pour 4 % des personnes enquêtées, on retrouve des salaires dans les fichiers fiscaux, alors qu'aucun salaire n'a été déclaré dans l'enquête. La proportion de personnes qui déclarent des salaires dans l'enquête mais pas auprès des impôts est un peu plus faible, de l'ordre de 3 %.

Tableau 14 : Déclarer ou non des salaires en fonction de la source

Enquête avant imputation	Enquête après imputation	DGFIP	Effectif brut	Effectif pond	% pond
pas de salaires	pas de salaires	pas de salaires	4 704	19 233 260	43%
salaires	salaires	salaires	5 077	20 595 288	46%
pas de salaires	pas de salaires	salaires	454	1 766 278	4%
pas de salaires ou salaire nul	salaires	pas de salaires	61	274 048	1 %
pas de salaires ou salaire nul	salaires	salaires	344	1 548 498	3 %
salaires	salaires	pas de salaires	295	1 037 123	2 %
			10 935	44 454 495	100 %

Champ : SRCV 2005 - DGFIP, individus ayant pu être appariés.

Lors de l'enquête, il est demandé à l'enquêté de se référer à sa déclaration d'impôt. L'enquêté peut refuser ou déclarer ne pas avoir rempli de déclaration. Les erreurs de déclaration sont plus importantes pour les individus qui n'ont pas consulté leur déclaration d'impôts pour répondre à l'enquête. 6 % (contre 3 %) d'entre eux ont déclaré ne pas avoir de salaires lors de l'enquête alors que des salaires ont été retrouvés dans les fichiers fiscaux. A contrario, 4 % (contre 2 %) ont déclaré avoir des salaires lors de l'enquête mais aucun salaire n'a été retrouvé dans les fichiers administratifs.

Tableau 15 : Déclarer ou non des salaires en fonction de la source et de la consultation de la déclaration pour répondre à l'enquête

Consultation de la déclaration d'impôts pour répondre						Pas de consultation de la déclaration d'impôts					
Enq brut	Enq imp	DGFIP	Effectif brut	Effectif pondéré	Pourcentage pondéré	Enq brut	Enq imp	DGFIP	Effectif brut	Effectif pondéré	Pourcentage pondéré
NON	NON	NON	2 940	12 603 140	38 %	NON	NON	NON	1 764	6 630 119	60 %
OUI	OUI	OUI	4 474	18 070 324	54 %	OUI	OUI	OUI	603	2 524 964	23 %
NON	NON	OUI	268	1 101 498	3 %	NON	NON	OUI	186	664 780	6 %
NON	OUI	NON	33	148 263	0 %	NON	OUI	NON	28	125 785	1 %
NON	OUI	OUI	166	779 236	2 %	NON	OUI	OUI	178	769 262	7 %
OUI	OUI	NON	163	663 399	2 %	OUI	OUI	NON	132	373 723	3 %
Total			8 044	33 365 860	100 %	Total			2 891	11 088 633	100 %

NON = pas de salaires ou salaire nul

OUI = salaires

Les personnes qui ne consultent pas leur déclaration d'impôts, ont plus tendance à déclarer des salaires dans l'enquête qu'aux impôts. Il peut s'agir d'un effet de mémoire, les revenus sont considérés comme des salaires alors qu'ils sont d'une autre nature, mais il peut également s'agir de revenus dissimulés à l'administration fiscale.

Les réponses par proxy sont moins précises que les réponses données directement par la personne concernée. Les hommes ont également tendance à moins bien répondre que les femmes. Ce résultat reste vrai « toute chose égale par ailleurs » bien que les hommes soient moins souvent présent que les femmes pour répondre (proxy) et qu'ils aient plus souvent que les femmes des revenus d'indépendants.

Les erreurs sont fréquentes pour les indépendants. Ils ont, plus souvent que la moyenne, tendance à faire des erreurs dans leurs types de revenus. Certains se considèrent comme indépendant mais touchent des salaires, d'autres nous déclarent des salaires alors qu'il s'agit de bénéfices d'indépendants.

Dans certains ménages, les erreurs proviennent d'une inversion entre conjoints. Les revenus de la femme sont déclarés pour l'homme et vice-versa, d'où un problème d'adéquation avec les données fiscales. Ces erreurs pourront être corrigées grâce à l'appariement entre les données d'enquête et les données fiscales à partir de l'enquête SRCV 2008.

Dernière catégorie à risque, les retraités. Ce résultat avait déjà été observé en comparant les résultats de l'enquête SRCV à ceux de l'enquête ERF. Les retraités oublient de déclarer leurs salaires lors de

l'enquête. Il s'agit en général de petits montants qui sont bien déclarés auprès de l'administration fiscale mais qui sont, soit oubliés, soit annexés aux retraites dans l'enquête.

Tableau 16 : Modélisation logistique de la concordance des sources

	Probabilité accord DGFIP enquête	Rapport des risques (Odd Ratio)
Constante	2,3543	
Non-consultation de la déclaration d'impôts pour répondre à l'enquête	-0,0786	0,854
<i>Consultation de la déclaration d'impôts</i>	<i>Ref.</i>	<i>Ref.</i>
Réponse personnelle	0,2172	1,544
<i>Réponse par proxy</i>	<i>Ref.</i>	<i>Ref.</i>
Homme	-0,1066	0,808
<i>Femme</i>	<i>Ref.</i>	<i>Ref.</i>
Célibataire	-0,3728	0,607
Marié	0,2469	1,129
<i>Veuf ou Divorcé</i>	<i>Ref.</i>	<i>Ref.</i>
Indépendant	-0,0401	0,923
Emploi à temps plein	0,5774	2,575
Emploi à temps partiel	0,3472	2,046
Chômeur	-0,4136	0,956
Élève, étudiant	-0,0454	1,382
Retraité	-0,0971	1,312
<i>Autre inactif</i>	<i>Ref.</i>	<i>Ref.</i>
Agriculteur, artisan	-1,4894	0,239
Cadre	0,6859	2,104
Profession intermédiaire	0,5090	1,763
Employé	-0,2250	0,846
Ouvrier	0,1909	1,283
Retraité	0,3863	1,559
<i>Inactif</i>	<i>Ref.</i>	<i>Ref.</i>

Lecture : ont été reportés ici les coefficients estimés des variables explicatives modèle logistique. Un coefficient de signe positif (resp. négatif), statistiquement significatif, indique que l'on est en présence d'un facteur qui accroît (resp. réduit) les chances, de se trouver dans telle situation plutôt que telle autre.

Tous les autres coefficients sont significatifs au seuil de 1 %

Ref. : catégorie de référence.

Champ : SRCV 2005 - DGFIP, individus ayant pu être appariés.

Tableau 17 : Paramètres estimés du modèle logit polytomique non ordonné opposant les trois situations suivantes : avoir déclaré des salaires aux impôts et dans l'enquête ou dans aucun des deux, avoir déclaré des salaires aux impôts mais pas dans l'enquête, avoir déclaré des salaires dans l'enquête mais pas aux impôts.

	Avoir déclaré des salaires aux impôts mais pas dans l'enquête <i>vs pas de contraction entre les 2 sources</i>	Avoir déclaré des salaires dans l'enquête mais pas aux impôts <i>vs pas de contraction entre les 2 sources</i>	Avoir déclaré des salaires aux impôts mais pas dans l'enquête <i>vs avoir déclaré des salaires dans l'enquête mais pas aux impôts</i>
Constante	-3,1332	-3,3998	0,2666
Non-consultation de la déclaration d'impôts pour répondre à l'enquête consultation de la déclaration d'impôts	-0,0115 <i>Ref.</i>	0,1769 <i>Ref.</i>	-0,1884 <i>Ref.</i>
Réponse personnelle Réponse par proxy	-0,0761 <i>Ref.</i>	-0,3543 <i>Ref.</i>	0,2782 <i>Ref.</i>
Homme Femme	0,0991 <i>Ref.</i>	0,1127 <i>Ref.</i>	-0,0136 <i>Ref.</i>
Célibataire Marié Veuf ou Divorcé	0,0252 -0,1167 <i>Ref.</i>	0,7739 -0,2550 <i>Ref.</i>	-0,7487 0,1383 <i>Ref.</i>
Emploi à temps plein Emploi à temps partiel Chômeur Élève, étudiant Retraité Autre inactif	-0,7513 -1,2790 0,9913 0,4023 0,1064 <i>Ref.</i>	-0,4154 0,1031 0,0243 0,3908 0,0326 <i>Ref.</i>	-0,3360 -1,3822 0,9670 0,0115** 0,0738 <i>Ref.</i>
Agriculteur, artisan Cadre Profession intermédiaire Employé Ouvrier Retraité Inactif	1,9331 -0,4653 -0,6799 -0,4400 -0,8235 -0,0852 <i>Ref.</i>	1,1550 -0,6042 -0,1150 0,8085 0,3660 -0,6918 <i>Ref.</i>	0,7781 0,1389 -0,5649 -1,2485 -1,1896 0,6066 <i>Ref.</i>

Lecture : ont été reportés ici les coefficients estimés des variables explicatives dans chacune des trois équations du modèle. Un coefficient de signe positif (resp. négatif), statistiquement significatif, indique que l'on est en présence d'un facteur qui accroît (resp. réduit) les chances, de se trouver dans telle situation plutôt que telle autre.

** significatif au seuil de 2 % ; Tous les autres coefficients sont significatifs au seuil de 1 %

Ref. : catégorie de référence.

Champ : SRCV 2005 - DGFIP, individus ayant pu être appariés.

3.2.2.2. Les montants déclarés

Les écarts dans les montants de salaires déclarés sont assez faibles. Si on considère les personnes qui ont déclaré un salaire dans les deux sources, l'écart¹⁹ moyen est de 74 euros en faveur de l'enquête et l'écart médian est de 0,5 euros en faveur de la DGFIP. Pour 80 % des individus, l'écart entre le salaire déclaré dans l'enquête et le salaire déclaré auprès de l'administration fiscale est négligeable (inférieur à une centaine d'euros par mois). En tête et en queue de distribution, les écarts peuvent être très importants. Pour 2 % des individus, l'écart est supérieur à 1 000 par mois.

Tableau 18 : Ecart entre le montant de salaire dans l'enquête et celui déclaré aux impôts

inférieur à -30 %	-30 % à -10 %	-10 % à 10 %	10 % à 30 %	supérieur à +30 %	Total
6,2 %	7,3 %	74,6 %	5,6 %	6,3 %	100 %

¹⁹ Salaire annuel dans l'enquête après imputation - Salaire annuel DGFIP.

Tableau 19 : Distributions des écarts entre le montant de salaire dans l'enquête et celui déclaré aux impôts

5 396 individus	Ecart individuel entre le salaire annuel dans l'enquête après imputation et le salaire annuel DGFIP
Min	-128 901
1 %	-19 091
5 %	-5 061
10 %	-2 202
20 %	-483
Q1	-82
30 %	-1
Médiane	-0.5
Q3	0
80 %	225
90 %	2 127
95 %	5 275
99 %	17 459
Max	157 398
Moyenne	74

Champ : SRCV 2005 - DGFIP, individus qui ont à la fois des salaires après imputation des données d'enquête et dans les fichiers fiscaux.

Tableau 20 : Distributions des écarts entre le montant de salaire en fonction de la consultation ou non de la déclaration pour répondre

Ecart individuel entre le salaire annuel dans l'enquête après imputation et le salaire annuel DGFIP	4626 Individus déclaration d'impôts pour répondre	770 Individus pas de déclaration d'impôts pour répondre
Min	-128 901	-39 626
1 %	-13 374	-24 235
5 %	-3 907	-9 262
10 %	-1 545	-5 967
20 %	-194	-2 583
Q1	-5	-1 902
30 %	0	-1 408
40 %	0	-438
Médiane	0	0
60 %	0	169
70 %	0	904
Q3	0	1 625
80 %	1	2 705
90 %	1 450	6 665
95 %	3 909	11 919
99 %	14 388	41 987
Max	157 398	150 852
Moyenne	-1	505

En moyenne, les écarts entre les salaires déclarés dans l'enquête et ceux déclarés auprès de l'administration fiscale sont moins importants pour les individus ayant accepté de consulter leur déclaration d'impôts pour répondre. Pour les trois quarts des individus, l'écart entre les deux sources est négligeable. Pour 12 % des individus l'écart entre les deux sources est supérieur à 30 %.

La consultation de la déclaration fiscale pour répondre réduit les écarts en moyenne mais elle ne permet pas d'éviter les gros écarts dus à des erreurs. Pour les individus ayant accepté de consulter leur déclaration de revenus, les écarts moyen et médian sont nuls alors d'en cas de non recours à la déclaration de revenus l'écart moyen est de 505 euros (médiane=0). Par contre, l'écart maximal est de 157 398 euros pour les individus ayant consulté leur déclaration et de 150 852 pour les individus ayant refusé de la consulter.

92 % des individus pour lesquels l'écart entre les deux sources est inférieur à 10% ont consulté leur déclaration de revenus pour répondre (85 % pour l'ensemble des salariés).

Les réponses part proxy favorisent également les écarts importants. En effet, le taux de recours au proxy est de 25 % pour l'ensemble des salariés et 35 % pour les salariés pour lesquels l'écart entre les deux sources est supérieur à 30 %. Il y a symétrie entre les écarts positifs et les écarts négatifs. Les erreurs ont donc tendance à se compenser en moyenne.

Les chômeurs ont, plus souvent que la moyenne, tendance à déclarer dans l'enquête un salaire éloigné de plus de 30 % de ce qu'ils ont déclaré à l'administration. Il s'agit aussi bien de sur-estimations que de sous-estimations. Les étudiants et les inactifs en général ont également tendance à faire des erreurs importantes. Cependant, il s'agit de population avec de faibles salaires et une erreur de quelques euros peut représenter un pourcentage important du salaire.

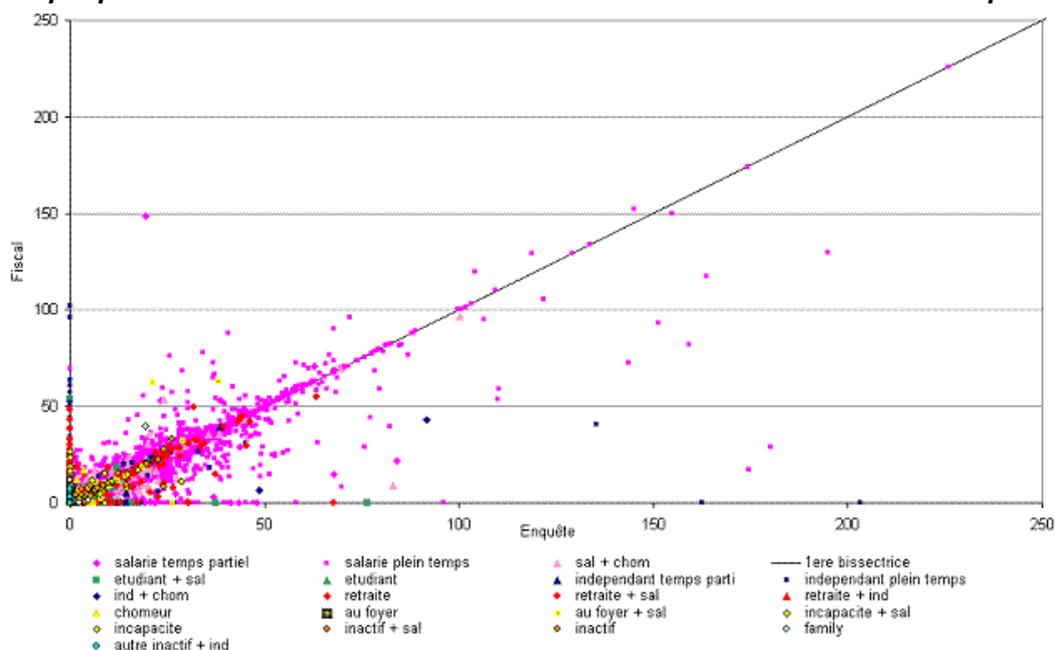
Tableau 21 : Ecart entre le montant de salaire déclaré dans l'enquête et montant déclaré aux impôts, en fonction du statut économique au moment de l'enquête

	Emploi à plein temps	Emploi à temps partiel	Chômeur	Élève, étudiant	Retraité	Autre inactif	Total
inférieur à -30 %	60,5	13,4	12,5	8,5	1,9	3,3	100
-30 % à -10 %	71,8	14,9	7,8	0,8	2,9	2,0	100
-10 % à 10 %	74,2	16,1	4,8	2,0	1,7	1,2	100
10 % à 30 %	76,5	12,8	5,2	2,6	2,6	0,3	100
supérieur à +30 %	58,2	17,6	13,0	3,5	4,7	3,1	100
Ensemble	72,3	15,7	6,0	2,5	2,0	1,5	100

Les écarts observés pour les salariés à temps plein sont plus souvent négligeables en pourcentage mais peuvent être importants en valeur brute.

Les écarts les plus importants, en valeur, observés sur les salaires concernent les salariés à plein temps et les indépendants.

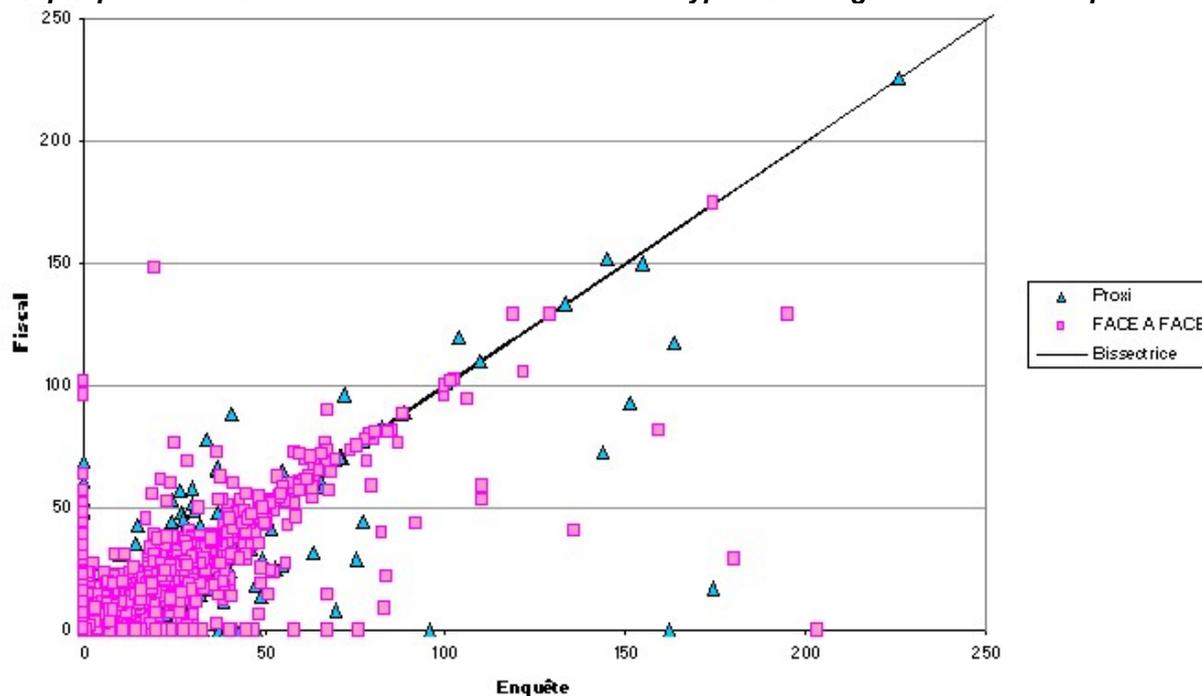
Graphique 5 : Salaires en fonction de la source et du statut déclaré dans l'enquête



Salaires annuels en milliers d'euros.

On a supprimé le point correspondant à la personne gagnant 900 000 euros corrigées à tort.

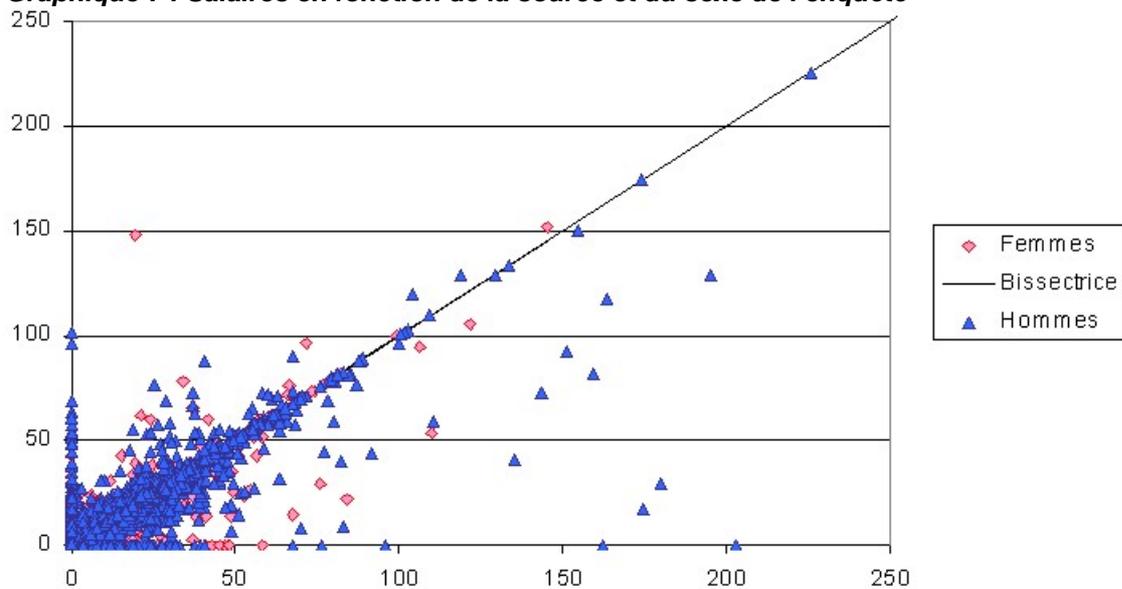
Graphique 6 : Salaires en fonction de la source et du type d'interrogation lors de l'enquête



Salaires annuels en milliers d'euros.

On a supprimé le point correspondant à la personne gagnant 900 000 euros corrigées à tort.

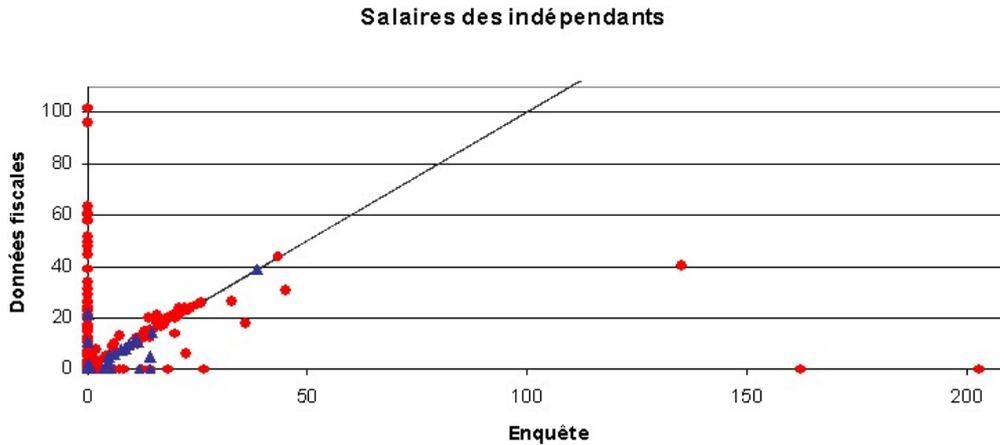
Graphique 7 : Salaires en fonction de la source et du sexe de l'enquêté



Salaires annuels en milliers d'euros.

On a supprimé le point correspondant à la personne gagnant 900 000 euros corrigées à tort.

Graphique 8 :

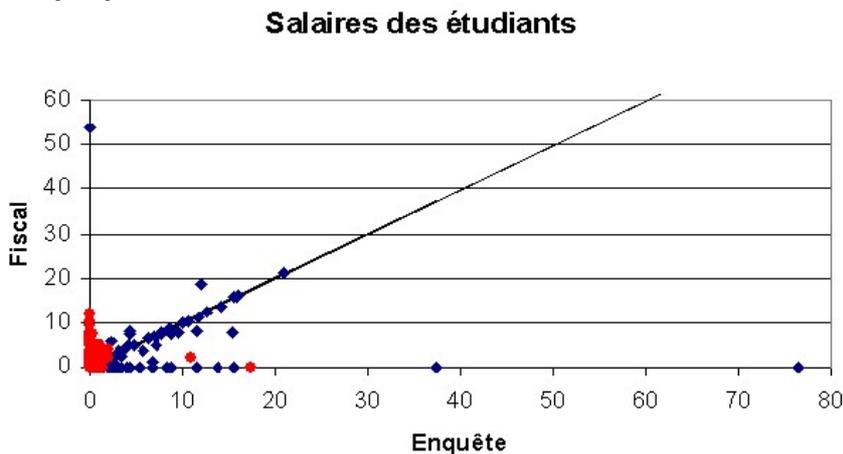


Salaires annuels en milliers d'euros.

En rouge, les indépendants à plein temps. En bleu, les indépendants à temps partiel.

Les valeurs les plus extrêmes correspondent à des confusions entre salaires et revenus d'indépendants. Les personnes déclarent des revenus d'indépendants dans l'enquête alors qu'en réalité elles touchent des salaires ou alors le contraire.

Graphique 9 :

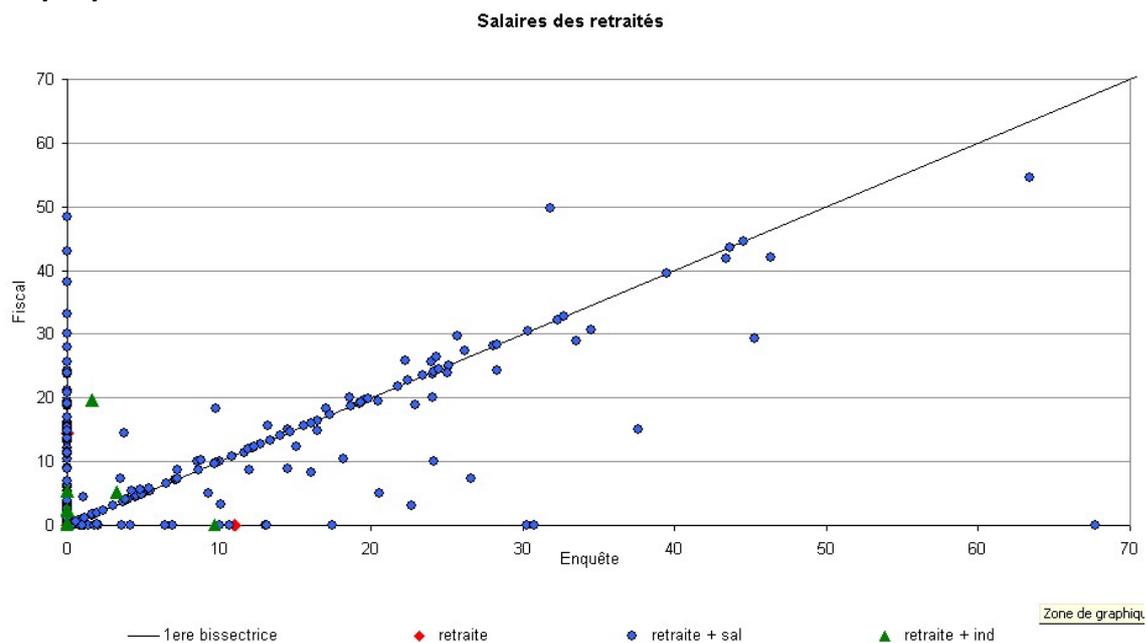


En milliers d'Euros

En bleu, ceux qui déclarent avoir un emploi salarié dans l'enquête

Les salaires des étudiants sont corrects quand ils sont déclarés mais les oublis sont fréquents, aussi bien dans l'enquête qu'auprès de l'administration fiscale.

Graphique 10 :

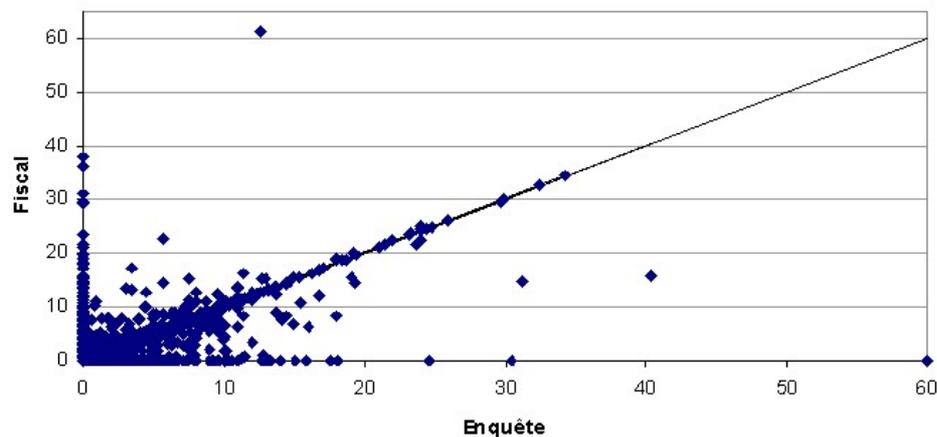


Une part importante des retraités oublie de déclarer ses salaires lors de l'enquête.

3.3. Impact sur les allocations chômage (PY090N)

Les indemnités de licenciement peuvent être comprises dans le montant de chômage donné dans l'enquête. Ces indemnités n'étant pas toujours imposables, elles peuvent être à l'origine des écarts en faveur de l'enquête. Ce cas est prévu dans la collecte 2008. Le montant des indemnités non imposables continue à être demandé à l'enquêté.

Graphique 11 : allocations chômage

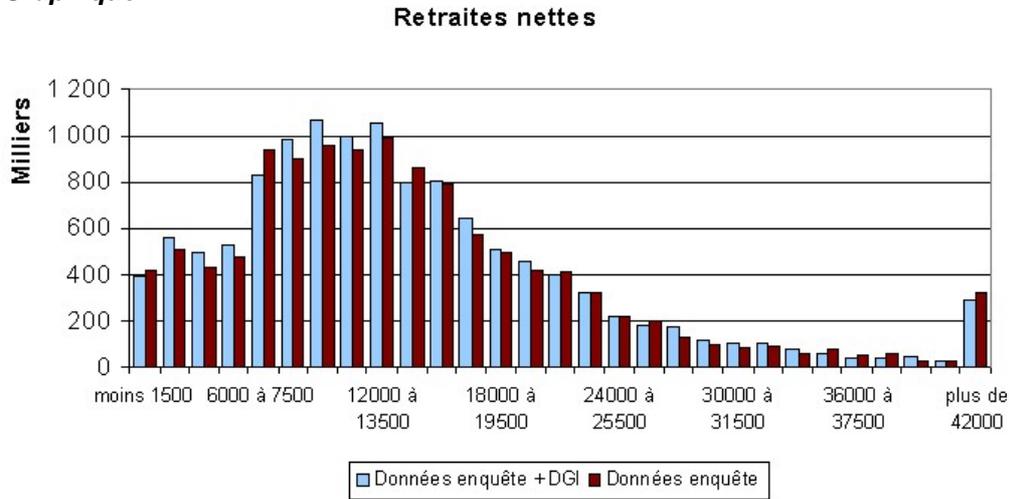


En milliers d'euros.

3.4. Impact sur les retraites (PY100N)

3.4.1. Distribution

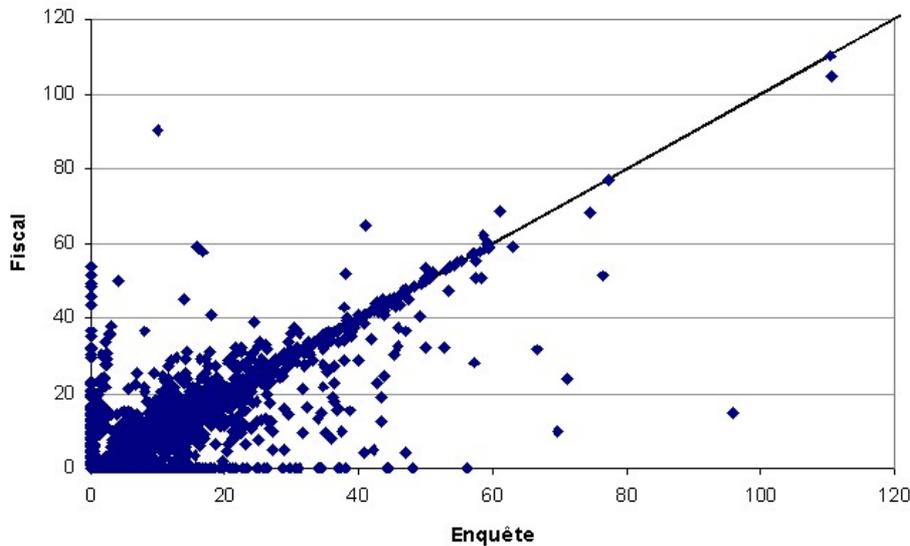
Graphique 12 :



L'impact de l'appariement sur les retraites est plus fort que sur les salaires. La forme de la distribution reste globalement la même mais on observe des décalages.

3.4.2. Impact individuel

Graphique 13 : Retraites en fonction de la source (zoom)



En milliers d'euros.

Les montants de retraites déclarés dans l'enquête sont supérieurs à ceux déclarés aux impôts. Deux phénomènes peuvent être à l'origine de ces écarts. Les retraités, qui ont tendance à oublier de déclarer leurs salaires dans l'enquête, compensent avec le montant de leur retraite. De nombreuses prestations pour les retraités ne sont pas imposables (majoration pour avoir élevé 3 enfants, retraite de combattant, rente défiscalisée, retraite exonérée). Ces montants sont en général inclus dans la variable PY100N mais ne sont pas dans le montant de retraite récupéré auprès de l'administration fiscale.

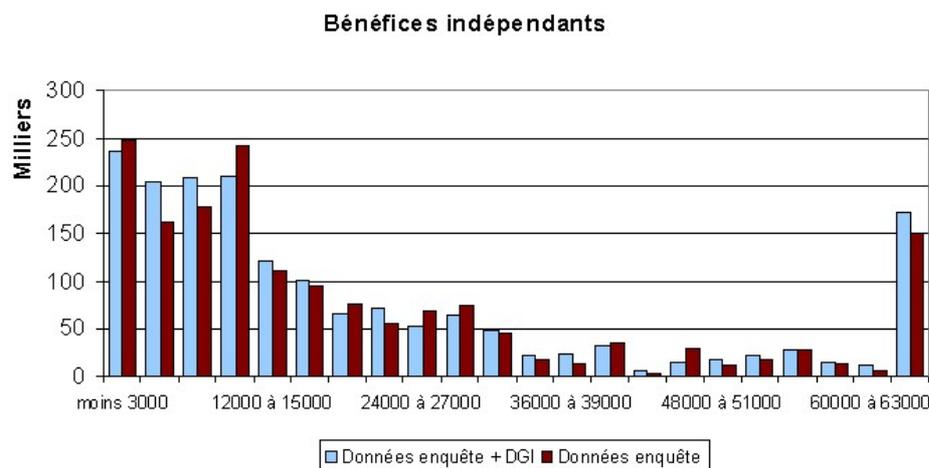
Ces montants non imposables sont demandés dans l'enquête 2008 et viendront compléter le montant récupéré par appariement.

3.5. Impact sur les bénéfices des indépendants (PY050N)

3.5.1. Distribution

C'est sur les bénéfices des indépendants que l'impact de l'appariement est le plus fort. En effet, cette information est difficile à collecter par voie d'enquête. Les ménages ne savent pas quoi déclarer à l'enquêteur. De plus les concepts ne sont pas exactement les mêmes aux impôts et dans l'enquête. Les bénéfices des indépendants sont difficiles à modéliser et donc à imputer et corriger quand nécessaire.

Graphique 14 :



3.6. Erreurs dans les types de revenus

3.6.1. Corrélations entre les écarts dans les revenus au niveau ménage

Il existe une corrélation entre les écarts dans les revenus déclarés et les écarts dans les impôts déclarés. Les ménages qui répondent de façon erronée pour leurs revenus ont également tendance à répondre de façon erronée sur le montant de leurs impôts. Ces ménages n'ont en général pas consulté leur feuille d'imposition pour répondre à l'enquête. Au niveau individuel, les corrélations entre les différents écarts sont très faibles. Peu de personnes perçoivent plusieurs sources de revenus. La corrélation la plus importante se trouve entre les salaires et les bénéfices des indépendants. En effet, la confusion entre ces deux sources de revenu est assez fréquente et explique les plus gros écarts observés sur ces types de revenus.

3.6.2. Comparaisons des types de revenus déclarés

Tableau 22 : Types de revenus déclarés dans l'enquête et auprès de l'administration pour les personnes déclarant ne pas avoir rempli de déclaration d'impôts

Types de revenus	Salaire	DGFIP	0	1	0	1	0	0	1	1	1	0	
		Enquête	0	0	1	1	0	0	0	0	0	1	0
	Indépendants	DGFIP	0	0	0	0	0	0	0	0	0	0	0
		Enquête	0	0	0	0	0	0	0	0	0	0	0
	Chômage	DGFIP	0	0	0	0	0	0	0	0	1	1	1
Enquête		0	0	0	0	0	0	0	0	0	0	0	
Retraite	DGFIP	0	0	0	0	1	1	1	0	0	0	0	
	Enquête	0	0	0	0	1	0	0	0	0	0	0	
Effectifs			1041	96	90	45	28	10	5	5	4	3	
			77,2%	7,1%	6,7%	3,3%	2,1%	0,7%	0,4%	0,4%	0,3%	0,2%	
enquête	revenu min		0	0	144	300	840	0	0				
	revenu moyen		0	0	7 042	10 384	7 432	0	0				
	revenu median		0	0	2 565	3 000	6 477	0	0				
	revenu max		0	0	96 000	72 000	19 236	0	0				
DGFIP	revenu min		0	17	0	187	348	129	3 787				
	revenu moyen		0	2 524	0	9 634	7 903	1 649	5 446				
	revenu median		0	1 569	0	3 025	6 403	1 450	4 125				
	revenu max		0	14 475	0	96 235	22 913	4 048	9 659				
Ecart	moyen		0	-2 524	7 042	750	-471	-1 649	-5 446				
	médian		0	-1 569	2 565	21	-6	-1 450	-4 125				

Lecture colonne 2 : 96 personnes (soit 7,1% des personnes qui déclarent ne pas avoir rempli de déclaration d'impôts) ont des salaires dans les fichiers fiscaux et aucun autre type de revenus, ni dans l'enquête, ni dans les fichiers fiscaux. Leurs salaires moyens dans les fichiers fiscaux sont de 2 524 euros annuels.

77 % des 1 349 individus qui déclarent ne pas avoir rempli de déclaration d'impôts n'ont aucun revenu, ni dans le fichier d'enquête, ni dans le fichier fiscal. 7 % ne déclarent aucun revenu dans l'enquête mais ont des salaires dans les fichiers fiscaux. Ces salaires sont très faibles, 2 524 euros annuels en moyenne et 14 475 euros au maximum.

90 individus (soit 7 % de ceux qui déclarent ne pas avoir fait de déclaration) déclarent des salaires dans l'enquête mais aucun revenus n'est retrouvé dans les fichiers fiscaux²⁰. Le salaire moyen déclaré par ces individus dans l'enquête est de 7 042 euros et atteint même 96 000 euros pour l'un d'entre eux. Il peut s'agir de travail dissimulé à l'administration. Les faibles effectifs incriminés ne permettent pas de faire une évaluation fiable des montants dissimulés à l'administration.

9 586 personnes déclarent lors de l'enquête avoir rempli une déclaration d'impôts. Pour près de 85 % d'entre eux, les types de revenus déclarés dans l'enquête et les types de revenus retrouvés dans les fichiers fiscaux coïncident. 44 % de ces 9 586 personnes ont uniquement des salaires dans les deux sources, 24 % ont uniquement des retraites, 5 % n'ont aucun revenu, 4 % ont des salaires et des allocations chômage, 3 % ont uniquement des revenus d'indépendants, 2 % ont uniquement des allocations chômage, 1 % ont des salaires et des retraites et moins de 1 % ont à la fois des salaires et des revenus d'indépendants. Pour ces personnes les montants déclarés sont très proches dans les deux sources. L'écart médian est toujours nul sauf pour les personnes qui ont uniquement des revenus d'indépendants mais l'écart s'explique par une différence de définition entre les deux sources. L'erreur la plus fréquente, 4 % des personnes, correspond à une absence de revenus dans l'enquête et à des salaires dans les fichiers fiscaux. Le salaire médian est alors de 15 700 euros. Viennent ensuite 2 % des personnes qui ont des retraites dans les fichiers fiscaux mais aucun revenu dans l'enquête.

Un peu plus de 1 % de ces personnes déclarent des salaires dans l'enquête mais aucun revenu dans les fichiers fiscaux. Les montants sont alors assez faibles, salaire annuel médian à 4 836 euros. Il peut

²⁰ L'individu n'est pas retrouvé dans les fichiers de la DGFIP.

s'agir de travail dissimulé mais il s'agit plus vraisemblablement de jeunes qui ont travaillé quelques mois l'été et qui n'ont pas déclaré ces revenus aux impôts.

1 % des personnes ont des salaires dans l'enquête et dans les fichiers fiscaux mais ont oublié les allocations chômage dans l'enquête. L'écart est alors faible. Les allocations chômage ont vraisemblablement été incluses dans les salaires lors de l'enquête.

Les retraités ayant oublié de déclarer leurs salaires dans l'enquête représentent 0,7 % des personnes interrogées déclarant avoir rempli une déclaration d'impôts. La même proportion a bien déclaré ses salaires dans l'enquête mais a oublié ses retraites.

Seules 0,4 % des personnes déclarent des revenus d'indépendants dans l'enquête et des salaires dans leur déclaration fiscale.

Les écarts importants observés au niveau individuel correspondent donc à des cas marginaux. Dans la majorité des cas, les types de revenus déclarés sont corrects.

Types de revenus déclarés dans l'enquête et auprès de l'administration pour les personnes déclarant avoir rempli une déclaration d'impôts

Types de revenus	Salaire	DGFIP	1	0	0	1	1	0	0	0	0	0	1	1	1	1	1	1	0	1	1	0	1	1	0	0	0	1		
		Enquête	1	0	0	1	0	0	0	0	0	1	1	1	0	1	0	1	0	0	0	1	0	0	1	0	0	0	1	
	Indépendants	DGFIP	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	1	0	
		Enquête	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	1	0	0	1	
	Chômage	DGFIP	0	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0
		Enquête	0	0	0	1	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0
Retraite	DGFIP	0	1	0	0	0	0	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
	Enquête	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	
Effectifs		4213	2269	514	392	378	298	178	156	135	96	76	65	65	62	53	42	39	38	36	29	29	29	25	23	22	21	17		
		43,9%	23,7%	5,4%	4,1%	3,9%	3,1%	1,9%	1,6%	1,4%	1,0%	0,8%	0,7%	0,7%	0,6%	0,6%	0,4%	0,4%	0,4%	0,4%	0,4%	0,3%	0,3%	0,3%	0,3%	0,2%	0,2%	0,2%	0,2%	
enquête	revenu min	40	113	0	840	0	100	45	0	96	967	2 222	1004	1160	450	1119	1524	713	658	52	199	500	6447	0	152	0	5200	7956		
	revenu moyen	20294	15380	0	13412	0	23602	9451	0	10211	15495	29013	17705	18682	5950	20153	27848	10641	18338	22528	9038	19203	52426	0	15534	0	13513	43920		
	revenu médian	17650	13012	0	11641	0	14550	8758	0	4836	12826	22442	14400	15156	4700	18186	20682	7320	12542	17996	6000	18000	33762	0	9000	0	12109	34184		
	revenu max	225791	280000	0	113013	0	260000	40342	0	67713	73925	165181	58214	151155	31200	85360	89500	48000	137784	65004	60000	45000	181884	0	72000	0	30000	165427		
DGI	Revenu min	6	200	0	840	15	-33492	297	50	0	2172	3036	1851	3027	1450	36	369	0	125	-6100	0	5160	297	2294	0	521	-8977	2244		
	revenu moyen	20243	15338	0	12935	16959	29773	9501	10491	0	16669	27168	35348	22592	12940	15904	25378	0	14277	24681	0	32759	33762	13211	0	6092	13024	23511		
	revenu médian	17765	13356	0	11503	15735	14260	9100	8345	0	12930	22095	19018	19299	9959	14550	20006	0	14346	18834	0	24182	24475	11249	0	5046	11399	22976		
	revenu max	225791	110235	0	97526	87998	441135	34306	51121	0	74368	147907	961576	146902	115077	69652	96126	0	33077	70090	0	101740	116025	41020	0	21100	32766	70754		
Ecart	Moyen	51	41	0	477	-16959	-6171	-50	-10491	10211	-1174	1845	-17643	-3909	-6989	4249	2469	10641	4061	-2152	9038	-13555	18664	-13211	15534	-13211	489	20409		
	Médian	-1	0	0	0	-15735	-1006	0	-8345	4836	-253	0	-765	-2858	-4296	2552	673	7320	474	-245	6000	-5165	0	-11249	9000	-11249	-549	9520		

3.7. Impact sur les impôts

3.7.1. L'impôt total (HY145N²¹)

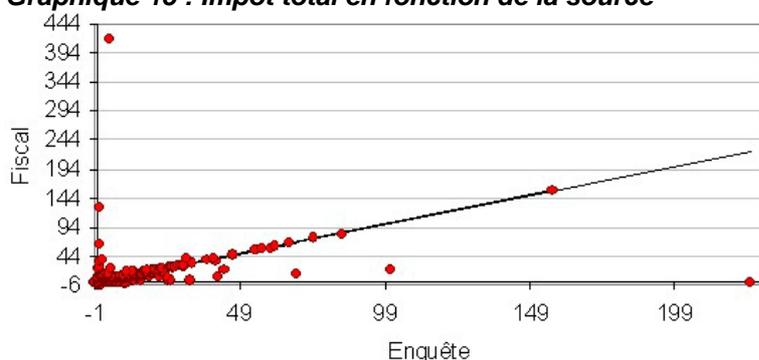
Tableau 23 : Distribution de l'impôt total en fonction de la source

	Nb ménages	Moyenne	Médiane	Min	Q1	Q3	Max
HY145_DGFIP	5 613	1 787	566	-6 471	0	2 031	417 356
HY145N	5 613	1 852	640	-983	70	2 077	225 554

Les montants d'impôts collectés par enquête et ceux collectés auprès de la direction générale des impôts sont très proches en moyenne. L'impôt moyen est de 1 787 euros en utilisant les fichiers de la DGFIP et de 1 852 euros dans l'enquête, soit une différence de 3,5 %.

L'impôt total se décompose en impôt sur le revenu, moins prime pour l'emploi, plus taxe d'habitation, plus impôts payés à l'étranger. Le montant de l'impôt payé à l'étranger provient uniquement de l'enquête. Les trois autres catégories d'impôts peuvent faire l'objet de comparaisons spécifiques.

Graphique 15 : Impôt total en fonction de la source



Impôts en milliers d'euros.

Les impôts sont, en général, bien déclarés dans l'enquête. Les écarts importants observés sont en faveur des données fiscales. Les ménages ont donc tendance à sous-estimer leurs revenus dans l'enquête mais à sur-estimer leurs impôts.

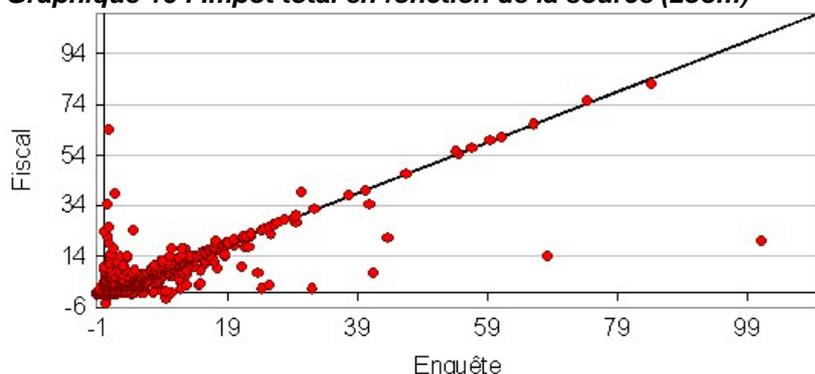
On peut lister quelques cas très particuliers :

Un ménage a déclaré 225 000 euros de taxe d'habitation dans l'enquête alors qu'il a très peu de revenus et que d'après le fichier de l'administration fiscale, il paie 200 euros de taxe d'habitation. Pour ce ménage, l'écart entre les deux sources est très important. Il s'agit vraisemblablement d'une erreur de saisie au moment de l'enquête.

Un ménage a un impôt sur le revenu de 400 000 euros mais le montant dans le fichier d'enquête a été mis à 0 car jugé trop élevé et donc correspondant à une erreur (il s'agit du ménage dont l'un des membre à un salaire de 900 000 euros).

Ces deux cas se compensent. Dans un cas une valeur a été corrigée à tort et dans l'autre, la valeur aberrante n'a pas été repérée lors des corrections.

Graphique 16 : Impôt total en fonction de la source (zoom)



Impôts en milliers d'euros.

²¹ HY145N = IRPP + TH - PPE + Impot_etranger

3.7.2. L'impôt sur le revenu (IRPP)

Tableau 24 : Distribution de l'impôt sur le revenu en fonction de la source

	Moyenne	Médiane	Min	Q1	Q3	Max
IRPP_DGFIP	1 492	312	-6 905	0	1 582	417 356
IRPP_enq	1 406	218	0	0	1 510	99 999

L'écart est plus important pour l'impôt sur le revenu. L'impôt sur le revenu moyen est de 1 492 euros dans les fichiers de la DGFIP et de 1 406 euros dans l'enquête, soit un écart de 6 %.

Les plus gros écarts sont dus à des erreurs Francs / Euros non corrigées où à des erreurs lors des imputations.

En général, le montant de l'impôt sur le revenu est bien renseigné par les ménages. Les plus gros écarts proviennent de montant non ou mal corrigés lors des imputations ou alors de montants volontairement trop élevés donnés par les ménages et non repérés en apurement.

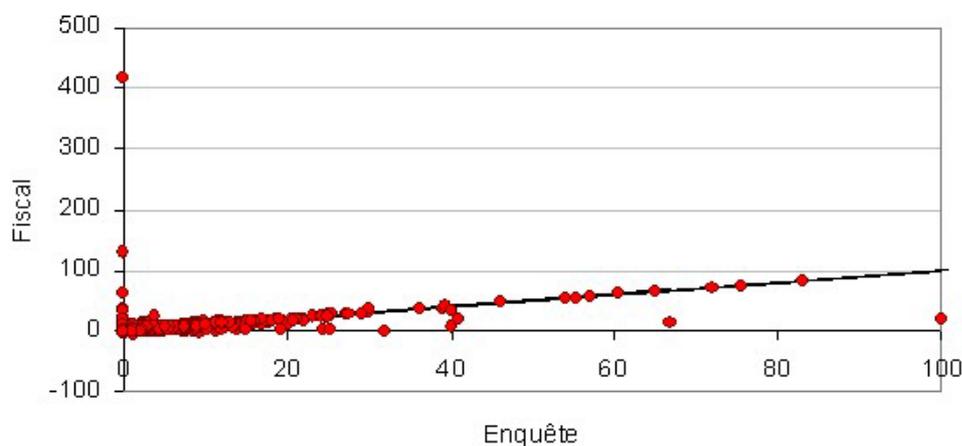
Un peu plus de 1 % des ménages ne donne pas de montant pour leur impôt sur le revenu lors de l'enquête. Pour 90 % d'entre eux, ce montant est récupéré lors de l'appariement.

14 % des ménages qui ont déclaré ne pas payer d'impôts sur le revenu lors de l'enquête ont un montant d'impôts positif dans les fichiers fiscaux.

Tableau 25 : Montant de l'impôt sur le revenu en fonction de la source

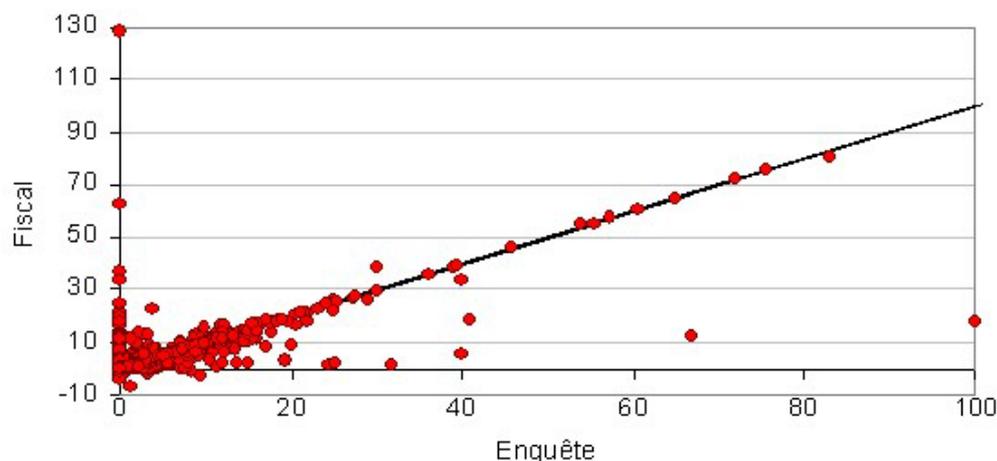
Enquête	DGFIP				Ensemble
	Valeur manquante	Impôt négatif	Impôt nul	Impôt positif	
Valeur manquante	9,2 %	1,3 %	29,4 %	60,1 %	1,3 %
Impôt nul	2,9 %	8,9 %	73,7 %	14,5 %	43,0 %
Impôt positif	1,2 %	1,3 %	3,2 %	94,3%	55,7 %
Ensemble	2,1 %	4,5 %	33,8 %	59,6 %	100 %

Graphique 17 : Impôt sur le revenu en fonction de la source



Impôts en milliers d'euros.

Graphique 18 : Impôt sur le revenu en fonction de la source (zoom)



Impôts en milliers d'euros.

3.7.3. La prime pour l'emploi (PPE)

D'après le fichier de la DGFIP, 28 % des ménages perçoivent la prime pour l'emploi. Seuls 10 % des ménages ont un montant non nul pour la prime pour l'emploi dans le fichier de l'enquête SRCV. Il est prévu dans le questionnaire que ce montant soit directement déduit du montant déclaré pour l'IRPP, ce qui explique en partie que le montant de l'impôt sur le revenu soit plus faible dans l'enquête que dans les données fiscales.

Tableau 26 : Distribution de la prime pour l'emploi en fonction de la source

	Moyenne	Médiane	Min	Q1	Q3	Max
PPE_DGFIP	92	0	0	0	80	1 456
PPE_enq	29	0	0	0	0	1 500

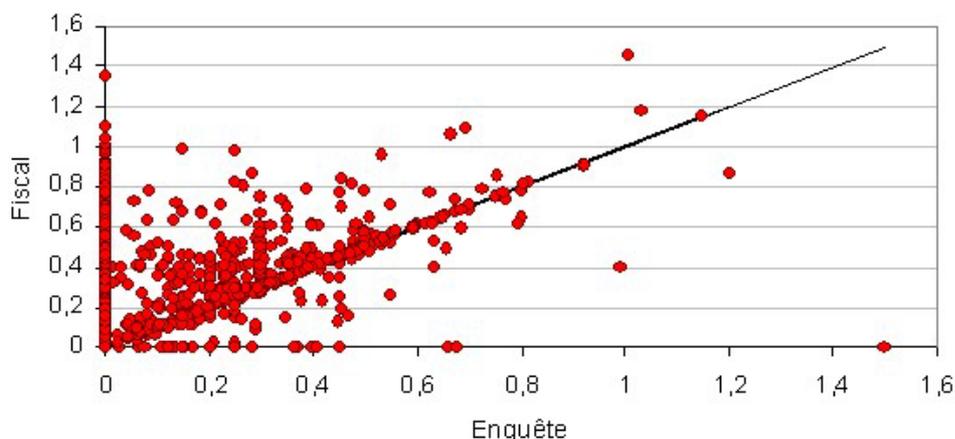
Il est donc plus pertinent de comparer le montant de l'impôt sur le revenu moins la prime pour l'emploi.

Tableau 27 : Distribution de l'impôt sur le revenu diminué de la prime pour l'emploi en fonction de la source

	Moyenne	Médiane	Min	Q1	Q3	Max
DGFIP	1 399	180	-6 905	0	1 543	417 356
Enq	1 377	189	-1 200	0	1 500	99 999

Même en tenant compte de la prime pour l'emploi, le montant de l'impôt déclaré dans l'enquête est plus faible que le montant collecté par l'administration fiscale mais l'écart est beaucoup plus faible (1,6 %) que lorsque l'on compare les valeurs brutes.

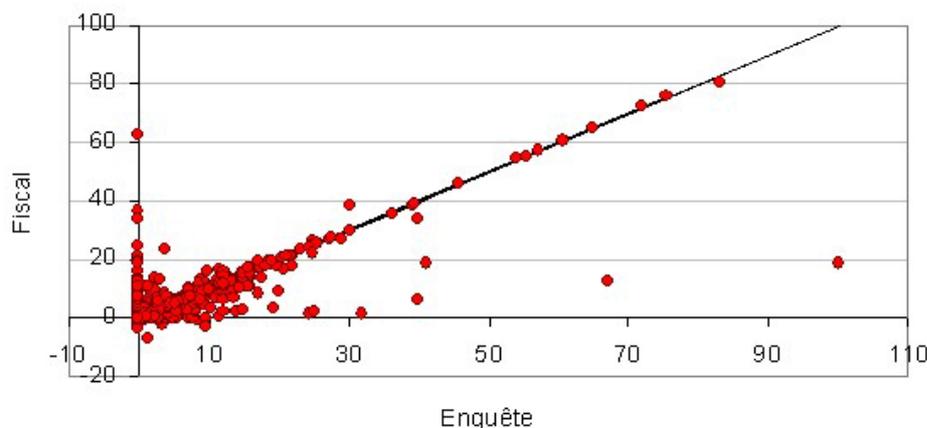
Graphique 19 : Prime pour l'emploi en fonction de la source



En milliers d'euros.

Les écarts concernant la prime pour l'emploi sont très importants. Les personnes enquêtées la déduisent en général directement du montant de leur impôt lors de l'enquête.

Graphique 20 : Impôt sur le revenu moins prime pour l'emploi en fonction de la source



En milliers d'euros.

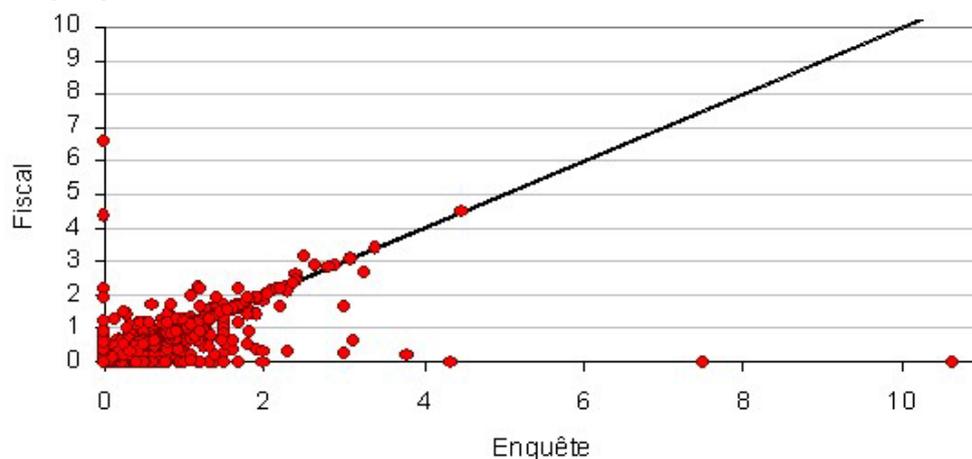
La valeur extrême pour l'IRPP a été supprimée.

3.7.4. La taxe d'habitation (TH)

Tableau 28 : Distribution de la taxe d'habitation en fonction de la source

	Moyenne	Médiane	Min	Q1	Q3	Max
TH_DGFiP	367	314	0	68	524	6 598
TH_enq	426	318	0	75	532	225 345

Graphique 21 : Taxe d'habitation en fonction de la source



En milliers d'euros.

La valeur extrême pour la TH a été supprimée.

Les écarts les plus élevés proviennent là encore de montants excessifs donnés lors de l'enquête et non corrigés en apurement. Les ménages confondent parfois taxe d'habitation et taxe foncière. Dans le cas où ils en ignorent les montants, ils ont la possibilité d'estimer la somme des deux. Ce montant est ensuite divisé par deux lors du traitement des données, ce qui peut donner lieu à des écarts plus ou moins importants.

4. Impact sur le suivi longitudinal

4.1. Les changements de niveau de vie

Tableau 29 : Changements de niveau de vie entre 2004 et 2005 croisés avec les changements entre 2005 et 2006 en utilisant les données d'enquête en 2005

2005 données d'enquête		2005/2006						
		< -30%	-30%< < -10%	-10%< < 10%	10%< < 30%	> 30%	mq	Total
2004/2005	< -30%	6%	7%	19%	15%	39%	13%	100%
	-30%< < -10%	5%	11%	31%	25%	19%	8%	100%
	-10%< < 10%	3%	10%	54%	17%	9%	6%	100%
	10%< < 30%	9%	20%	42%	14%	8%	7%	100%
	> 30%	23%	17%	32%	11%	9%	9%	100%
	mq	8%	13%	38%	16%	15%	9%	100%
	Total	9%	13%	41%	16%	13%	8%	100%

Lecture : En utilisant les données d'enquête pour 2004, 2005 et 2006, 6 % des individus qui ont vu leur niveau de vie baisser de plus de 30 % entre 2004 et 2005 ont connu la même baisse entre 2005 et 2006.

Tableau 30 : Changements de niveau de vie entre 2004 et 2005 croisés avec les changements entre 2005 et 2006 en utilisant les données fiscales en 2005

2005 données DGI		2005/2006						
		< -30%	-30%< < -10%	-10%< < 10%	10%< < 30%	> 30%	mq	Total
2004/2005	< -30%	7%	7%	17%	10%	47%	12%	100%
	-30%< < -10%	5%	10%	30%	25%	22%	8%	100%
	-10%< < 10%	3%	11%	55%	17%	8%	6%	100%
	10%< < 30%	8%	20%	42%	16%	8%	7%	100%
	> 30%	20%	19%	29%	12%	10%	9%	100%
	mq	9%	13%	39%	15%	15%	9%	100%
	Total	8%	13%	40%	16%	15%	8%	100%

Lecture : En utilisant les données d'enquête pour 2004 et 2006 et les données fiscales pour 2005, 7 % des individus qui ont vu leur niveau de vie baisser de plus de 30 % entre 2004 et 2005 ont connu la même baisse entre 2005 et 2006.

Les évolutions de niveau de vie calculées en utilisant les données fiscales sont globalement les mêmes que celles calculées en utilisant uniquement les données d'enquête.

Le fait d'utiliser les données fiscales ne provoque pas de changement, en moyenne, dans les évolutions de niveau de vie.

Tableau 31 : Changements de niveau de vie entre 2004 et 2005 en fonction de la source utilisée en 2005

		dgi 2004/2005						
		< -30%	-30%< < -10%	-10%< < 10%	10%< < 30%	> 30%	mq	Total
enq 2004/2005	< -30%	59%	17%	11%	5%	7%	0%	100%
	-30%< < -10%	15%	54%	20%	5%	6%	0%	100%
	-10%< < 10%	5%	9%	73%	8%	5%	0%	100%
	10%< < 30%	5%	6%	24%	51%	14%	0%	100%
	> 30%	6%	4%	11%	14%	65%	0%	100%
	mq	0%	0%	0%	0%	0%	100%	100%
	Total	8%	11%	29%	12%	13%	27%	100%

Lecture : 59 % des individus qui ont connu une baisse de plus de 30 % de leur niveau de vie calculé avec les données d'enquête entre 2004 et 2005, ont connu la même baisse si on calcule leur niveau de vie avec les données d'enquête 2004 et les données fiscales 2005.

Tableau 32 : Changements de niveau de vie entre 2005 et 2006 en fonction de la source utilisée en 2005

		dgi 2005/2006						mq	Total
		< -30%	-30% < -10%	-10% < 10%	10% < 30%	> 30%			
enq 2005/2006	< -30%	54%	20%	17%	5%	5%	0%	100%	
	-30% < -10%	10%	53%	24%	8%	5%	0%	100%	
	-10% < 10%	3%	7%	74%	9%	7%	0%	100%	
	10% < 30%	3%	3%	21%	55%	18%	0%	100%	
	> 30%	5%	7%	15%	12%	60%	0%	100%	
	mq	0%	0%	0%	0%	0%	100%	100%	
Total		8%	13%	40%	16%	15%	8%	100%	

Lecture : 54 % des individus qui ont connu une baisse de plus de 30 % de leur niveau de vie calculé avec les données d'enquête entre 2005 et 2006, ont connu la même baisse si on calcule leur niveau de vie avec les données fiscales 2005 et les données d'enquête 2006.

74 % des individus gardent la même évolution de niveau de vie entre 2004 et 2005 quelle que soit la source pour les données 2005. Cette proportion est de 67 % entre 2005 et 2006. Les décalages, quand ils sont lieu, se font avec les catégories limitrophes. Peu de personnes changent radicalement d'évolution de niveau de vie en fonction de la source utilisée pour les données 2005. 12 % (resp. 14 %) des individus ne sont pas dans la même catégorie ni dans la catégorie limitrophe. Les plus gros décalages ont lieu en queue de distribution mais ils concernent peu de personnes.

4.2. Passage du seuil de pauvreté

Tableau 33 : Passages de seuil en fonction de la source de données 2005

2004	2005 données fiscales	2005 données enquête	2006	Pourcentage
Pas de changement				
PAUVRE	PAUVRE	PAUVRE	PAUVRE	75,1
NON PAUVRE	NON PAUVRE	NON PAUVRE	NON PAUVRE	4,1
Changement mais cohérence 2005				
NON PAUVRE	PAUVRE	PAUVRE	PAUVRE	3,6
PAUVRE	PAUVRE	PAUVRE	NON PAUVRE	3,3
PAUVRE	NON PAUVRE	NON PAUVRE	NON PAUVRE	1,4
NON PAUVRE	NON PAUVRE	NON PAUVRE	PAUVRE	1,2
PAUVRE	NON PAUVRE	NON PAUVRE	PAUVRE	0,9
NON PAUVRE	PAUVRE	PAUVRE	NON PAUVRE	0,8
Changement du à une incohérence 2005				
PAUVRE	NON PAUVRE	PAUVRE	PAUVRE	3,3
NON PAUVRE	PAUVRE	NON PAUVRE	NON PAUVRE	0,9
Changement décalé d'un an en fonction de 2005				
PAUVRE	PAUVRE	NON PAUVRE	PAUVRE	1,8
NON PAUVRE	NON PAUVRE	PAUVRE	NON PAUVRE	0,6
PAUVRE	PAUVRE	NON PAUVRE	NON PAUVRE	0,9
PAUVRE	NON PAUVRE	PAUVRE	NON PAUVRE	0,9
NON PAUVRE	PAUVRE	NON PAUVRE	PAUVRE	0,7
NON PAUVRE	NON PAUVRE	PAUVRE	PAUVRE	0,6

79 % des personnes ne changent pas de statut au cours des 3 années considérées, quelle que soit la source de données pour 2005. 11 % changent de statut au cours de ces 3 années mais les résultats sont cohérents pour l'année 2005 quelle que soit la source d'information. Dans 90 % des cas, le fait d'avoir recours aux données administratives ne modifie pas le statut de la personne face à la pauvreté.

Dans 4 % des cas, la personne garde le même statut au cours des 3 années au regard des données d'enquête mais change en 2005 au regard des données fiscales. Ces cas peuvent être considérés comme les plus graves. Pour ces personnes, l'impact de l'appariement est fort.

Dans 2 % des cas, la personne ne change pas de statut si on considère les données fiscales mais change de statut en 2005 si on considère les données d'enquête.

Dans les 4 % des cas restant, le changement de statut est retardé ou avancé d'un an en fonction de la source d'information utilisée en 2005.

5. Impact sur le questionnaire 2008

5.1. Une simplification du questionnement

Un des objectifs de la refonte est d'alléger le questionnement de l'enquête SRCV. En effet, grâce à l'appariement, on peut alléger sensiblement les questionnaires (ménage et individu). L'architecture du questionnaire reste inchangée, certaines questions sont supprimées.

On continue à questionner le ménage sur les types de revenus qu'il perçoit. La stratégie de balayage est maintenue, seules les indicatrices sur la nature des revenus perçus sont conservées. Certains montants ne sont plus demandés. On ne demande plus le montant des salaires, des allocations chômage, des informations figurant sur la déclaration fiscale pour les indépendants et les retraités. Par contre, on continue à demander les montants qui ne figurent pas sur la déclaration fiscale : revenus immobiliers et financiers, salaires non imposables, retraites exonérées.

Il n'est plus demandé au ménage d'avoir recours à ses documents pour répondre. La refonte du dispositif permet donc de gagner du temps lors de la collecte, à la fois en allégeant le questionnaire et en allégeant le temps de recherche de l'information par le ménage.

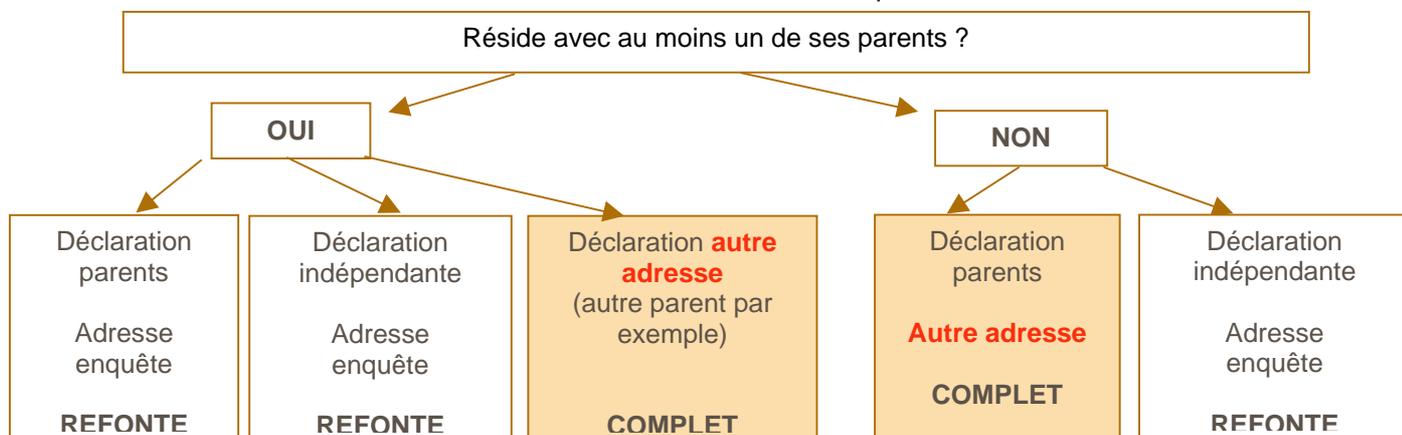
5.2. Le cas particulier des jeunes

Les jeunes entre 18 et 25 ans peuvent choisir de déclarer leurs revenus avec les parents. Cette disposition pose des difficultés pour l'appariement dans le cas où le jeune n'est pas interrogé à l'adresse où il a fait sa déclaration.

Une disposition particulière est donc mise en place dans le questionnement SRCV.

Pour les personnes entre 16 et 25 ans, la partie « revenus » du questionnaire individuel commence par un questionnement sur leur déclaration de revenus. Dans le cas où leurs parents sont encore en vie, on leur demande s'ils ont fait leur déclaration avec leurs parents et on vérifie s'ils habitent ou non avec eux (ou lui).

S'ils font leur déclaration séparément de leurs parents mais à l'adresse de l'enquête ou s'ils font leur déclaration avec leurs parents et vivent avec eux, il n'y a pas de problème pour l'appariement. Le questionnement est alors allégé comme pour tous les adultes de plus de 25 ans. S'ils font leur déclaration à une autre adresse que celle de l'enquête, l'appariement ne sera pas possible. Le questionnement est alors le même qu'avant la refonte. Tous les montants de revenus seront demandés et ils devront sortir leur déclaration de revenus ou leur récapitulatif de salaires.



6. Conclusion

La collecte des revenus imposables par appariement avec les fichiers fiscaux à la place de la collecte par enquête permet de diminuer le temps d'enquête et allège la charge pour les ménages. Les personnes interrogées n'ont plus besoin de chercher leurs documents administratifs pour pouvoir répondre à l'enquête ce qui est très apprécié.

De plus l'appariement permet de collecter des informations plus fiables sur les revenus imposables, en particulier les salaires. L'effet proxy devrait diminuer, voire disparaître pour ces variables.

Pour les revenus qui ne sont pas intégralement imposables des questions complémentaires continuent à être posées lors de l'enquête. De même, le questionnement sur les types de revenus perçus est maintenu afin de pouvoir effectuer les redressements nécessaires en cas de besoin.

Le cas des jeunes qui ne vivent pas à l'adresse où ils déclarent leurs revenus a également été prévu dans le questionnaire SRCV 2008. Leurs revenus continueront à être intégralement demandés lors de l'enquête.

L'appariement devrait avoir peu d'impact sur les indicateurs macro : taux de pauvreté, revenu disponible, seuil de pauvreté...

Au niveau individuel, l'impact peut être très fort, mais dans des cas peu nombreux. Ils sont en général dus à des erreurs lors de l'enquête (Francs/Euros, un 0 de trop ...) ou alors à des valeurs extrêmes corrigées à tort. En moyenne, au niveau macro, ces deux types d'erreurs se compensent. Ces erreurs sont assez facilement repérables et pourront être contrôlées pour permettre un suivi longitudinal cohérent pour ces individus.

Les écarts plus faibles sont plus difficiles à expliquer et à repérer. Ils sont néanmoins nombreux. 10 % des individus se situent de part et d'autre du seuil de pauvreté suivant la source de données que l'on utilise. Là encore les individus qui passent au-dessus du seuil compensent ceux qui passent en dessous mais il faudra veiller à s'assurer d'un suivi cohérent au niveau individuel.

Bibliographie

Lollivier S. et Verger D., 2005, Trois apports des données longitudinales à l'analyse de la pauvreté, *Économie et Statistique* n^{os} 383-384-385.

Lollivier S. et Verger D., 1987, Les comportements en matière d'épargne et de patrimoine, *Économie et Statistique* n^{os} 202.

Cécile Bourreau-Dubois, Olivier Guillot et Éliane Jankeliowitch-Laval, 2001, Le travail à temps partiel féminin et ses déterminants, *Économie et Statistique* nos 349-350.

Rapport intermédiaire sur la qualité des données françaises E.U.-S.I.L.C. 2005 (portant sur les revenus de 2004 et sur les conditions de vie en 2005) #Données transversales.

Sylvie Michaud, David Dolson, Donna Adams, Martin Renaud, 1995, Combinaison des données administratives et des données d'enquête en vue d'alléger le fardeau des répondants dans les enquêtes longitudinales, documents de recherche de l'EDTR n° 75F0002M, Août 1995.

Alain Desrosières, 2004, Enquêtes versus registres administratifs : réflexions sur la dualité des sources statistiques, *Courrier des Statistiques* - 111 - septembre 2004.

Mylène Chaleix et Stéfan Lollivier, 2005, Des panels pour les statistiques sociales, *Courrier des Statistiques* - 113-114 - juin 2005.

CONSEIL NATIONAL DE L'INFORMATION STATISTIQUE, 2006, Enquêtes statistiques et sources administratives : une complémentarité à exploiter, *Chroniques* n° 5 - Décembre 2006.

CONSEIL NATIONAL DE L'INFORMATION STATISTIQUE, Mars 2007, Rapport du groupe de travail Niveaux de vie et inégalités sociales, Président : Jacques Freyssinet, Rapporteurs : Pascal Chevalier, Michel Dollé.

Cédric AFSA ESSAFI, Les modèles logit polytomiques non ordonnés : théorie et applications, INSEE, Série des Documents de Travail Méthodologie Statistique N° 0301.