

Sources et méthodes

Source Revenus disponibles localisés (RDL)

Le dispositif sur les « Revenus disponibles localisés » (RDL) vient compléter ceux existant concernant l'information sur les revenus des ménages (l'enquête nationale Revenus fiscaux et sociaux et le dispositif Revenus fiscaux localisés).

Son principal apport réside dans le fait de disposer de données sur les revenus disponibles (c'est à dire y compris les prestations sociales et après déduction des Impôts directs et de la CSG et CRDS), les niveaux de vie et la pauvreté monétaire au niveau départemental.

Les deux autres sources sur les revenus des ménages sont accessibles dans cette rubrique « *Sources et méthodes* » aux emplacements suivants : enquête Revenus fiscaux et sociaux et dispositif Revenus fiscaux localisés.

Présentation générale

La demande sociale pour une meilleure connaissance des revenus disponibles et de la pauvreté monétaire au niveau départemental est très forte. Celle-ci est régulièrement exprimée par les acteurs publics régionaux, ainsi que par le conseil national de l'information statistique dans le cadre de ses travaux sur les niveaux de vie et les inégalités (voir le site du Cnis, Rapport N°103, Niveaux de vie et inégalités sociales - mars 2007 ; accessible par la rubrique *Publications* puis *Rapport*).

L'Insee diffuse les revenus disponibles des ménages au niveau national à partir de l'enquête Revenus fiscaux et sociaux (ERFS) et les revenus déclarés (ou revenus fiscaux) des ménages au niveau local à partir de la source Revenus fiscaux localisés (RFL).

La méthodologie d'élaboration de ces deux sources statistiques, complémentaires au dispositif Revenus disponibles localisés décrit ici, est accessible dans cette rubrique « *Sources et méthodes* » aux emplacements suivants :

- Enquête Revenus fiscaux et sociaux (ERFS) ;
- Revenus fiscaux localisés (RFL).

Les objectifs du dispositif Revenus disponibles localisés (RDL)

Le dispositif Revenus disponibles localisés (RDL) permet le calcul de revenus disponibles, au niveau départemental, sur le champ des « ménages fiscaux ». L'Insee met ainsi à disposition un ensemble d'indicateurs sur les revenus après redistribution comparables à ceux déjà publiés au niveau national. Les données de base de la source RDL sont celles du revenu déclaré par les foyers français à l'administration fiscale (dite source fiscale).

Son insertion parmi les autres sources sur les revenus des ménages

L'exploitation de la source RDL a été réalisée pour la première fois à partir des revenus de l'année 2004 déclarés en 2005 et de la taxe d'habitation au 1^{er} janvier 2005. Le dispositif RDL a traité ainsi, cette année là, les données fiscales concernant plus de 23 millions de ménages fiscaux.

Les différents impôts payés par les ménages sont connus par la source fiscale. Les prestations sociales sont imputées, selon des méthodes les plus proches possibles de celles utilisées depuis plusieurs années au niveau national pour l'enquête Revenus fiscaux et sociaux (ERFS). Les imputations sont effectuées au niveau des départements ; le nombre d'allocataires et les montants imputés sont ensuite confrontés aux données fournies par les caisses d'allocations familiales (CAF). Les résultats entre les deux sources sont généralement très proches et lorsque ce n'est pas le cas (pour le RSA socle, par exemple), un tirage est effectué au niveau départemental parmi des allocataires potentiels, de façon à se rapprocher le plus possible des données des CAF¹.

Enfin, un ensemble de contrôles est mené pour s'assurer de la cohérence avec l'enquête nationale Revenus fiscaux et sociaux. La recherche de cohérence avec les déciles de niveaux de vie et le taux de pauvreté diffusés au niveau national a conduit à mettre en place des procédures de calage.

¹ Les deux sources ayant des champs un peu différents, les données issues de RDL n'ont pas de raison d'être rigoureusement calées sur celles des CAF.

Le dispositif RDL complète la source Revenus fiscaux localisés (RFL) pour ce qui est de la mesure des revenus déclarés évalués à un échelon local (le département). En outre, seul le dispositif RFL permet une exploitation locale à un niveau géographique infra-départemental des revenus déclarés. RDL ne remplace pas non plus, au niveau national, l'enquête Revenus fiscaux et sociaux pour la mesure des revenus disponibles et des indicateurs associés (niveau de vie, taux de pauvreté...). L'enquête Revenus fiscaux et sociaux est en effet plus complète : les estimations de prestations y sont bien plus précises car elles mobilisent des données réelles collectées auprès de la CNAF ; l'appariement avec l'enquête Emploi permet en outre de disposer d'une large palette d'informations sociodémographiques utiles pour caractériser les ménages ou individus dits « pauvres ».

Les données fournies à l'Insee

Pour élaborer la source Revenus disponibles localisés (RDL), l'Insee dispose des deux types de fichiers suivants, transmis chaque année par la direction générale des finances publiques (DGFIP) :

- Le fichier des **déclarations de revenus** (dit fichier POTE)

Ce fichier, pour l'année N, contient les données relatives aux déclarations des revenus de l'année N (le formulaire n°2042) déposées par les contribuables en mai N+1. Avec ce type de données, on répertorie des foyers fiscaux. Utilisé seul, ce fichier permet seulement de travailler par foyer fiscal et par personne (en dénombrant les personnes inscrites sur une même déclaration fiscale).

Chaque foyer fiscal y est repéré à l'aide d'un identifiant valable au sein de la direction fiscale dont il dépend. En général, la direction fiscale correspond au département sauf pour les départements très peuplés des Bouches du Rhône, du Nord et des Hauts-de-Seine divisés en deux directions fiscales, et Paris, divisé en cinq directions fiscales.

- Le fichier de la **taxe d'habitation** (dit fichier PLFC)

Le fichier de la taxe d'habitation comptabilise les logements pour lesquels les occupants sont susceptibles de payer la taxe d'habitation (TH), les personnes reliées à ces logements et leurs caractéristiques ainsi que les foyers fiscaux associés. L'Insee dispose de l'état de ce fichier au 1er janvier de l'année N+1. Le fichier recense des locaux de différentes natures : logements taxés au titre de résidence principale ou secondaire, autres locaux taxés séparément tels que parkings, garages ou caves. On en extrait les observations correspondant aux personnes reliées à des logements taxés en résidence principale.

Dans ce fichier, la personne redevable de la taxe d'habitation est identifiée d'une part par un identifiant de logement, qui permet de reconstruire le ménage fiscal, et d'autre part par un identifiant de foyer fiscal identique à celui du fichier des déclarations de revenus. Ceci permet de passer de la notion de foyer fiscal à celle de ménage fiscal : tous les foyers fiscaux rattachés au même logement sont regroupés sous le même identifiant de logement (celui du redevable de la taxe d'habitation) pour constituer le ménage fiscal.

Les fichiers exhaustifs cités précédemment contiennent des informations nominatives sensibles qui se doivent d'être protégées en vertu des recommandations de la Commission nationale de l'informatique et des libertés (Cnil). Les deux grandes opérations de production statistique utilisant cette source administrative sont donc regroupées à l'Insee au sein d'un seul et unique pôle, dont les personnels sont les seuls habilités à avoir accès aux fichiers.

Les millésimes : de la déclaration de revenus à la diffusion d'indicateurs statistiques

Entre l'année de référence des revenus et leur diffusion au niveau supra-communal, s'écoulent environ un an et demi. Ce délai provient du « circuit » emprunté par les données correspondantes. On peut résumer les étapes entre l'année N de référence des revenus et l'année N+2 de diffusion des résultats de la manière suivante.:

- ⇒ Les revenus de l'année N sont déclarés au début de l'année N+1 (en mai à compter de l'année 2005).
- ⇒ Au cours de l'année N+1 et au début de l'année N+2, les données sont traitées par les services fiscaux en vue du recouvrement de l'impôt sur le revenu (IR). Plusieurs « émissions » du fichier des déclarations de revenus sont nécessaires pour traiter la totalité des déclarations (notamment pour les travailleurs indépendants). Parallèlement, les services fiscaux chargés de la gestion de la taxe d'habitation procèdent à la mise à jour de leurs données de manière à fournir un fichier qui retrace la « campagne taxe d'habitation » de l'année N : le fichier transmis est « l'image » de la situation arrêtée au 1er janvier N+1.
- ⇒ Au début du premier trimestre N+2, les données commencent à être exploitées afin d'alimenter les trois sources statistiques qui en découlent :
 - l'enquête Revenus fiscaux et sociaux ;
 - le dispositif Revenus fiscaux localisés ;
 - le dispositif Revenus disponibles localisés.

Le champ du dispositif Revenus disponibles localisés (RDL)

Les informations utiles pour la source RDL sont donc directement extraites des fichiers de déclarations de revenus de la direction générale des finances publiques (DGFIP). Ces données sont appariées avec celles du fichier de la taxe d'habitation.

Le champ retenu est ainsi le même que celui de la source Revenus fiscaux localisés (RFL).

Ce champ se limite aux « ménages fiscaux » ; les communautés (foyers, hôpitaux, maisons de retraite) en sont donc exclues. Toutefois, par construction, toutes les personnes rattachées à la déclaration de revenus d'au moins un des membres du ménage fiscal sont incluses dans ce dernier. Ainsi, un étudiant qui dispose d'un logement indépendant ou une personne handicapée en institution peuvent être intégrés dans le ménage fiscal de leurs parents.

De par cette spécificité, les ménages de la source RDL ne correspondent pas aux ménages au sens du recensement de la population (RP), ni à ceux de l'enquête Revenus fiscaux et sociaux (et plus généralement à ceux au sens des enquêtes réalisées auprès des ménages), notamment pour le nombre de personnes dans ces ménages.

Par exemple, les ménages fiscaux sont ainsi un peu moins nombreux et de plus grande taille que les ménages au sens du recensement de la population.

Les ménages fiscaux constitués par des étudiants (occupant un logement indépendant et effectuant leur propre déclaration de revenus) sont par ailleurs pris en compte dans RDL.

Les limites de la source

Les données figurant dans les déclarations de revenus (dites déclarations n°2042) ne permettent pas de connaître tous les liens familiaux. En particulier, on ne sait pas déterminer avec certitude certains couples de concubins. On reconstitue ces couples de concubins à partir de critères de sexe et d'âge compatibles.

Cas particuliers

Le cas des enfants majeurs rattachés fiscalement à leurs parents

La législation fiscale autorise les parents à rattacher leurs enfants majeurs ou mariés sur leur propre déclaration de revenus s'ils sont âgés de moins de 21 ans quelle que soit leur situation, ou s'ils sont âgés de moins de 25 ans et poursuivent leurs études. Les enfants majeurs rattachés à la déclaration fiscale de leurs parents peuvent occuper un logement indépendant. Par défaut, ils sont pourtant inclus dans le ménage de leurs parents. En effet, la situation familiale décrite dans la déclaration de revenus des parents ne permet pas de repérer le fait que les enfants majeurs comptés fiscalement à charge occupent ou non le même logement que leurs parents. Par conséquent, si ces enfants majeurs ou mariés occupent un autre logement, leur inexistence en tant que foyer fiscal entraîne l'impossibilité de les créer en tant que ménage fiscal. Cette situation concerne essentiellement des étudiants.

Par comparaison avec le recensement de la population, ceci génère une sous-estimation du nombre de ménages et d'habitants des villes étudiantes et une surestimation de la taille des ménages des parents. En termes d'évaluation des niveaux de revenus des ménages, cela apparaît en revanche cohérent dans la mesure où ces étudiants sont effectivement à la charge de leur famille.

La situation fiscale des enfants majeurs décrite précédemment n'est pas la seule possible. En effet, les parents qui subviennent aux besoins de leurs enfants majeurs ont également la possibilité de ne plus les compter à charge, mais de déduire de leurs revenus les pensions alimentaires versées. Les enfants majeurs qui bénéficient de ces pensions sont alors tenus de les déclarer en leur nom. S'ils occupent de surcroît un logement indépendant, ils deviennent alors des « ménages fiscaux » à part entière.

Le cas des enfants et petits-enfants en résidence alternée

Dans le dispositif Revenus disponibles localisés (RDL), les enfants et petits-enfants en garde alternée sont pris en compte dans chacun des deux logements où ils résident. Ce traitement introduit des doubles comptes, à la différence de la source Revenus fiscaux localisés (RFL) qui elle ne prend pas en compte ces situations. Toutefois, ces cas de doubles comptes sont peu fréquents.

Le cas des communautés

Les personnes vivant en communautés (maisons de retraite, cités universitaires, foyers, communautés religieuses, centres d'hébergement, casernes, établissements militaires et autres structures collectives) constituent une population particulière, difficile à appréhender dans les fichiers fiscaux, et donc exclue des traitements et des statistiques produites.

En effet, le mode de gestion des communautés dans les fichiers de la taxe d'habitation (TH) est hétérogène entre les différentes directions de l'administration fiscale et il n'est de plus pas possible d'identifier ces structures. Si le fichier de la taxe d'habitation est tenu de les répertorier, elles peuvent être en revanche identifiées de multiples manières par les services fiscaux : un même type de communauté peut ainsi donner lieu à des codifications très variables.

En pratique, pour faire partie du champ, il faut que les foyers fiscaux soient présents à la fois dans le fichier de la taxe d'habitation et dans celui des déclarations de revenus. On élimine les habitations où l'on trouve plus de dix foyers fiscaux : en effet, étant donné le nombre élevé de foyers, on suppose qu'il s'agit d'une communauté : maison de retraite, cité universitaire, foyer, etc.

Les données sociodémographiques dans RDL

Les données sociodémographiques disponibles dans RDL sont celles que l'on trouve dans la déclaration de revenus. Ainsi on ne dispose pas, par exemple, de la catégorie socioprofessionnelle ni de l'âge exact des enfants, mais simplement de leur année de naissance. Au contraire, l'enquête Revenus fiscaux **et sociaux** permet de disposer de l'ensemble des informations socio-démographiques collectées dans le cadre de l'enquête Emploi.

En revanche, par rapport à la source Revenus fiscaux localisés (RFL), le dispositif Revenus disponibles localisés (RDL) intègre des indicateurs sur les tranches d'âge des individus, et non plus seulement sur l'âge du seul référént fiscal. Ceci permet de disposer d'informations, au niveau départemental, sur la pauvreté des enfants et des personnes âgées. Il est possible également à partir de RDL de caractériser les types de ménages (couples avec enfants, familles monoparentales et personnes seules, par exemple) et le type d'espace où sont situés les ménages considérés (catégorie d'aire urbaine et type de commune rurale ou urbaine à partir de RDL 2009).

Les types de ménages ou familles

A partir des foyers fiscaux identifiés avec les déclarations de revenus, il est possible de construire une typologie des ménages ou familles. Afin de pouvoir imputer des prestations sociales aux ménages, les familles (au sens des caisses d'allocations familiales) sont reconstruites dans le dispositif RDL. Les couples mariés ou pacsés sont clairement identifiés. Les liens de concubinages sont quelquefois connus, mais ce n'est pas toujours le cas. On reconstitue alors les couples de concubins à partir de critères « les plus courants » de sexe (deux personnes de sexe différent) et d'âge (la différence d'âge doit être inférieure ou égale à 15 ans pour que les deux personnes soient concubins). Avec ce second critère, l'objectif est d'éviter de considérer comme concubin les enfants du déclarant. Ainsi, lorsqu'un homme et une femme cohabitent dans le même logement, et lorsque la différence d'âge est inférieure à 15 ans, on considère que ces deux personnes vivent en concubinage et constituent une famille. On peut ainsi considérer comme concubins des personnes qui ne le sont pas (frère et sœur, par exemple). En revanche, on ne prend pas en compte les couples homosexuels non pacsés. Cette méthode est donc imparfaite, mais les informations disponibles ne permettent pas de mettre en œuvre une méthode plus élaborée.

Ensuite, pour pouvoir qualifier le type de ménage, on regarde si les personnes seules déclarant un revenu dans un ménage fiscal peuvent être les enfants d'un couple de référence du ménage ou de la personne la plus âgée. Lorsqu'il y a plusieurs couples ou que les différences d'âge ne permettent pas de considérer que des liens parents-enfants sont plausibles, le ménage est alors considéré comme un « ménage complexe ».

Les redressements effectués

Afin d'améliorer la précision de la source Revenus disponibles localisés (RDL), on utilise une méthode de redressement par « calage sur marges ». Cette méthode consiste à modifier les poids des ménages de telle sorte que le total de certaines variables, en l'occurrence celles citées ci-dessous, estimé à partir des données fiscales soit égal au « vrai » total (ou « marges ») connu par ailleurs (en l'occurrence par d'autres sources statistiques).

Un premier calage par type de ménage et décile de niveau de vie permet d'assurer la cohérence avec la structure de la population de l'Enquête nationale Revenus Fiscaux et Sociaux (ERFS).

Le second calage, au niveau individu, après attribution des revenus financiers non déclarés, permet d'assurer la cohérence de la distribution des niveaux de vie entre l'enquête Revenus fiscaux et sociaux (ERFS) qui fait référence au niveau national et le dispositif RDL. Le niveau de vie médian des individus comptabilisés dans la source RDL est calé sur le niveau de vie médian correspondant de l'enquête Revenus fiscaux et sociaux. Avant calage en effet, les deux distributions sont très proches, mais pas absolument identiques. Le poids des ménages les plus pauvres dans RDL, qui comprend une part importante des ménages étudiants, est ainsi plus important que dans ERFS. Cette dernière exclut ce type de ménages de son champ, considérant que les données fiscales ne permettent pas de disposer d'une mesure fiable de leur niveau de vie. Malgré les différences de champ avec ERFS, les déciles de niveaux de vie ne diffèrent pas de plus d'1 % entre les deux sources.

Après calage, le niveau de vie médian d'une année donnée et donc le seuil de pauvreté est identique dans les deux sources enquête Revenus fiscaux et sociaux et Revenus disponibles localisés. Il est cependant impossible d'assurer une cohérence complète entre les deux sources, le champ comme la composition des ménages étant en effet différents.

Les calages réalisés dans le dispositif Revenus disponibles localisés (RDL) introduisent donc une différence avec les revenus déclarés issus de la source Revenus fiscaux localisés (RFL) mais les distributions des revenus déclarés par département sont proches entre les deux sources.

D'une manière générale, pour des travaux portant spécifiquement sur une analyse locale (infra-départementale) des revenus, la source RFL reste la source de référence.

Quelques précautions d'analyse

Les comparaisons des données RDL 2011 avec celles diffusées précédemment doivent être menées avec beaucoup de précaution :

- ***elles sont limitées aux comparaisons depuis 2008,***
- ***elles sont limitées aux données relatives à l'ensemble de la population des zones étudiées.***

En effet, les données nationales ERFS sur lesquelles sont calées les résultats de RDL font régulièrement l'objet de rétopolations qui ne sont pas prises en compte dans le processus RDL. Dans ce contexte, les comparaisons depuis RDL2008 sont possibles car ces années sont sans changement majeur sur ERFS, mais elles sont exclues pour les années antérieures.

Les évolutions seront donc examinées avec grande précaution, entre zones et au regard de l'évolution constatée au niveau national. On veillera en particulier à ne pas interpréter trop finement des évolutions d'amplitude faible (à quelques dixièmes de points), du fait du degré d'imprécision lié à la méthodologie utilisée (imputations massives de prestations et revenus financiers puis calage sur l'enquête nationale ERFS).

De même, les évolutions des indicateurs par sous-population (taille et type de ménages, âge des individus) sont sensibles au degré d'imprécision lié à la méthodologie utilisée et ne doivent pas être interprétées. **Pour les sous populations, l'analyse en évolution est donc à proscrire, on se limitera ainsi à une analyse spatiale en niveaux relatifs sur une année donnée, en veillant à ne pas interpréter les écarts de faible ampleur.**