

Sources et méthodes

Enquête Revenus fiscaux (ERF)

Depuis 1996, les enquêtes Revenus fiscaux (ayant une périodicité annuelle) s'appuient sur les données de l'enquête Emploi de l'Insee (trimestrielle depuis 2003) appariées avec les données administratives émanant de la direction générale des Impôts (déclarations de revenus). La méthodologie d'élaboration de l'enquête, dans sa forme actuelle, est présentée ici. L'enquête d'une année donnée fournit une approche du revenu disponible monétaire des ménages pour l'année considérée, enrichie des informations de l'enquête Emploi collectées durant le 4ème trimestre de l'année.

L'objectif de cette enquête est de mesurer un revenu économique : ce dont les ménages ont disposé au cours d'une période pour consommer ou épargner. La richesse de l'enquête Revenus fiscaux réside également dans les différents concepts de revenus qu'elle permet de mesurer et ce afin d'évaluer le niveau de vie des ménages, la pauvreté monétaire, etc. Les différents concepts que l'enquête permet de mesurer sont exposés dans cette rubrique « *Sources et méthodes* » (voir la note : [les concepts de revenus de l'enquête Revenus fiscaux](#)).

Présentation générale de l'enquête

L'enquête Revenus fiscaux permet de produire des statistiques et de réaliser des études en matière de revenus, de niveau de vie, de pauvreté monétaire des [ménages](#) résidant en France métropolitaine et ce à l'aide d'indicateurs usuels d'analyse de la distribution des revenus (nombres, quartiles, déciles, moyenne, médiane, etc.) et aussi d'indicateurs de structure et de composition des ménages. Elle permet ainsi l'analyse des revenus suivant des critères sociodémographiques : [catégorie socioprofessionnelle](#) et âge des personnes composant le ménage, activité au sens du [BIT](#) de ces personnes, taille du ménage, etc. Elle précise également comment se cumulent les divers types de revenus (salaires, chômage, pensions, revenus agricoles, industriels, commerciaux, non commerciaux, etc.) perçus par chaque membre du ménage.

Il ne s'agit pas à proprement parler d'une enquête, puisqu'une partie des données recueillies, celles portant sur les revenus, n'est pas collectée par enquête mais par exploitation d'une source administrative.

L'enquête Revenus fiscaux se déroule en deux étapes.

La **première** consiste en un appariement statistique du fichier de l'[enquête Emploi](#) en continu (enquête trimestrielle depuis 2003), correspondant aux données de l'enquête du 4^{ème} trimestre de l'année N avec les fichiers fiscaux des revenus de l'année N. Le principe de l'appariement consiste à essayer de retrouver les déclarations fiscales des individus composant les ménages enquêtés à l'enquête Emploi. L'enquête Emploi en continu (EEC) est une enquête par sondage, c'est-à-dire que l'on interroge tous les trimestres un échantillon représentatif de l'ensemble de la population. Elle se déroule tout au long de l'année.

Dans un **second temps**, les informations sur les revenus non fournies par la source fiscale sont complétées par des estimations réalisées par l'Insee. Cette étape est indispensable pour passer du concept de revenu déclaré (ou revenu fiscal) à des concepts de revenus économiques (revenu disponible, niveau de vie, etc.). Par ailleurs, cette enquête, comme toutes les enquêtes par sondage, fait l'objet d'un redressement afin d'être représentative au niveau de la France métropolitaine.

Ainsi, l'enquête Revenus fiscaux d'une année donnée fournit une approche du revenu disponible monétaire des ménages pour l'année considérée, enrichie des données de l'enquête Emploi collectées durant le 4^{ème} trimestre de l'année.

Champ et limites de l'enquête

Le champ couvert par l'enquête est celui des individus des [ménages](#) dits « ordinaires » résidant en France métropolitaine. Sont donc exclus les personnes vivant en collectivité (maisons de retraite, communautés religieuses, les cités universitaires, les foyers de travailleurs, les prisons, etc.) et les personnes sans domicile.

L'enquête Revenus fiscaux est représentative des ménages métropolitains (approximativement 98 % de la population totale française hors [Dom-Com](#)) et elle a pour unité statistique le ménage (et non l'individu ou le [foyer fiscal](#)). Plus précisément, l'échantillon de l'enquête Revenus fiscaux *stricto sensu* est constitué des ménages répondants à l'enquête Emploi et pour lesquels on a retrouvé au moins une déclaration fiscale.

Par conception, l'enquête est tributaire des revenus partiellement ou non appréhendés par la source fiscale et de la législation fiscale qui peut faire varier le revenu déclaré d'une année sur l'autre.

Par ailleurs, les enquêtes Revenus fiscaux ne donnent pas d'information sur certains revenus non soumis à l'impôt, car par définition ils n'apparaissent pas sur le formulaire de déclaration de revenus

que chaque contribuable doit remplir (déclaration 2042). Ainsi les revenus du patrimoine, dont certains sont soumis au prélèvement libératoire, sont sous-estimés dans les enquêtes Revenus fiscaux. L'enquête ne collecterait que 12 % à 23 % des revenus des valeurs mobilières enregistrés par la comptabilité nationale et à peu près 50 % des revenus des patrimoines immobiliers ¹.

Il faut distinguer trois types de revenus du patrimoine :

- ✓ Ceux soumis au prélèvement libératoire : le prélèvement libératoire est une retenue forfaitaire versée directement à l'administration fiscale par l'établissement qui détient les placements. C'est le cas pour des placements à revenu fixe comme les revenus d'obligations ou les bons du Trésor.
- ✓ Ceux qui sont imposés au barème progressif de l'impôt sur le revenu et qui figurent donc sur la déclaration 2042 : c'est le cas des dividendes et des revenus des obligations pour les contribuables n'ayant pas opté pour le prélèvement libératoire, celui des revenus fonciers, ou de certains revenus ne pouvant être qualifiés de revenus professionnels et relevant de la catégorie des bénéfices industriels et commerciaux (BIC), des bénéfices non commerciaux (BNC) ou des revenus accessoires.
- ✓ Ceux qui sont exonérés d'impôt : c'est le cas des livrets d'épargne A, de l'épargne logement, des plans d'épargne en actions (PEA), de l'assurance-vie,...

Les évolutions de législation fiscale peuvent avoir une incidence sur le revenu déclaré (ou revenu fiscal) ; c'est par exemple le cas pour les revenus des indépendants (extension des régimes micro-entreprise par exemple).

L'enquête Revenus fiscaux de l'année N reconstitue le revenu disponible au titre des revenus déclarés pour l'année N et non le revenu disponible réellement perçu sur la période :

Par exemple :

- ✓ *Dans l'enquête Revenus fiscaux de l'année N, le revenu disponible du ménage est égal au revenu déclaré en N, diminué de l'impôt sur le revenu de l'année N, augmenté des prestations relatives à la législation en vigueur l'année N, qui sont estimées en fonction du revenu déclaré en N et de la composition du ménage observée au 4ème trimestre N (l'âge des enfants étant toutefois recalculé au 1er juillet N).*
- ✓ *Dans la réalité, le revenu disponible réel du ménage de l'année N est égal au revenu déclaré en N, diminué de l'impôt sur le revenu de l'année N-1 (c'est celui qui est payé en N), augmenté des prestations versées en N, qui sont calculées par les caisses d'allocation familiale (CAF) en fonction des revenus déclarés en N-1 (ou N-2) et de la composition du ménage tout au long de l'année N (avec l'âge des enfants au cours de l'année N).*

Ce décalage entre ce qui est reconstitué dans les enquêtes Revenus fiscaux et le revenu disponible réel n'a guère d'incidence pour les ménages dont les caractéristiques (en termes de revenus, situation familiale, etc.) sont stables au cours du temps.

Au niveau macroéconomique, pour le calcul des prestations sous conditions de ressources, le fait de retenir les revenus de l'année N peut conduire à sous estimer ou sur estimer les effectifs d'allocataires et/ou les montants des allocations estimés dans l'enquête.

¹ Voir à ce sujet : « [Les revenus du patrimoine dans les enquêtes revenus fiscaux](#) », document de travail, n°F0404, Insee, 2004.

Les évolutions de la méthode d'enquête

A l'origine quadriennales, les enquêtes Revenus fiscaux sont reconduites chaque année depuis 1996 grâce à un processus d'appariement statistique mis en place entre les données de l'enquête Emploi (annuelle au départ et trimestrielle depuis 2003) et celles des fichiers fiscaux.

Les premières enquêtes Revenus fiscaux (1956, 1962, 1965, 1970, 1975, 1979, 1984 et 1990) : assises sur les recensements de la population

A l'époque, l'Insee tirait un échantillon de [logements](#) dans l'échantillon maître du [recensement de la population](#). Pour chacun de ces logements, un document (dénommé « fiche adresse ») comportant les coordonnées géographiques du logement ainsi que la composition du ménage (quand elle était connue) était rempli. Il était ensuite transmis aux services fiscaux qui prenaient en charge la recherche des données dans les dossiers fiscaux des contribuables. L'Insee traitait ensuite les informations recueillies.

Une fois les questionnaires renseignés par les agents de l'administration fiscale, toutes les mentions nominatives ayant servi à l'identification des ménages étaient éliminées ; des questionnaires anonymes étaient donc retournés à l'Insee. La commission nationale informatique et libertés (CNIL) n'a pas autorisé l'Insee à reconduire cette procédure pour les enquêtes suivantes. Une nouvelle méthode a donc été mise au point et a reçu un avis favorable de la CNIL. Cette méthode repose sur un appariement entre fichiers informatiques, à la charge de l'Insee.

Les enquêtes depuis 1996 : mobilisation de l'enquête Emploi et appariement avec les déclarations de revenus

Sur la période 1996-2001, les données des enquêtes Revenus fiscaux (ERF) provenaient de celles de l'enquête Emploi annuelle de mars N+1 que l'on appariait avec le fichier des déclarations fiscales de l'année de revenu N. On récupérait ainsi, s'ils existaient, les revenus de l'ensemble des individus présents dans le ménage de l'enquête Emploi. Les ménages dont certains individus n'avaient pas été retrouvés dans la ou les déclarations fiscales font partie de la « non-réponse partielle », et des revenus leur étaient donc imputés. L'échantillon obtenu après imputation des revenus manquants étaient redressé par [calage sur les marges](#) de l'enquête Emploi et sur les masses fiscales de revenu. Pour arriver à un niveau de vie des ménages, des imputations de prestations familiales, minima sociaux et d'allocations de logement étaient réalisées. Ces imputations de transferts sociaux étaient effectuées sur barèmes ou, parfois, à l'aide d'équations économétriques (cas des allocations de logement attribuées aux propriétaires accédants). Les données obtenues respectaient, dans la mesure du possible, les statistiques fournies par les CAF des allocataires au 31/12 de l'année de revenu (année N) de l'enquête.

Depuis cette deuxième série d'enquêtes, la démarche des enquêtes Revenus fiscaux est globalement restée la même jusqu'à aujourd'hui, bien que la méthode d'élaboration ait été substantiellement modifiée en raison du changement important introduit par le passage d'une enquête Emploi annuelle à une enquête dite « en continu » (trimestrielle). Les enquêtes s'appuient donc sur un échantillon représentatif de « ménages ordinaires » de France métropolitaine, issu de l'enquête emploi, et leurs déclarations fiscales correspondantes. Les revenus fiscaux sont ensuite complétés par une estimation des revenus sociaux.

Les principales modifications récentes dans la méthodologie ont été les suivantes. Tout d'abord, en 2002, la méthodologie de l'enquête a été revue en raison du passage de l'[enquête Emploi annuelle](#) à

[l'enquête Emploi en continu](#). De ce fait, certaines informations disponibles jusque-là dans l'enquête Revenus fiscaux ne le sont plus, notamment celles concernant le calendrier d'activité des personnes interrogées. Puis, à compter de 2003, les revenus soumis à prélèvement libératoire figurant sur les déclarations fiscales sont intégrés au revenu disponible des ménages, ce qui constitue une amélioration de la couverture des revenus du patrimoine. Par ailleurs, en 2003, l'enquête a pris en compte les données du recensement rénové de la population. Ces modifications ont généré une rupture de série dans les données publiées. Une version rétropolée de l'enquête Revenus fiscaux pour l'année 2002 a ainsi été produite, de manière à assurer le passage d'une série à l'autre et donc la comparabilité dans le temps.

La taille de l'échantillon a également fluctué au cours du temps. De l'ordre de 23 000 ménages en 1996, il est passé à 45 000 en 1997, puis 70 000 à partir de 1998. Depuis 2002, avec le passage à l'enquête emploi en continu, l'échantillon a été ramené à environ 35 000 ménages.

Méthodologie et déroulement de l'enquête (en vigueur depuis 2002)

L'échantillon de départ

L'enquête Emploi en continu (EEC), socle de l'enquête Revenus fiscaux actuelle, est une enquête par sondage, c'est-à-dire que l'on interroge tous les trimestres un échantillon représentatif de l'ensemble de la population. Elle se déroule tout au long de l'année (d'où son appellation « en continu »).

L'échantillon est construit à partir du recensement général de la population de 1999 qui est une base exhaustive de logements calibrée pour être utilisable jusqu'en 2010 ; on tire des adresses de logements, parmi des aires géographiques situées en France métropolitaine et constituées de 20 logements en moyenne (on parle d'échantillon aréolaire). L'échantillon est renouvelé par sixième tous les trimestres. En première et sixième vague, l'interrogation des individus des logements se déroule en vis-à-vis. Pour les autres vagues, les enquêtes se déroulent par téléphone. Au bout d'un an, les 2/3 des logements de l'échantillon ont été remplacés.

Les logements ainsi tirés entrent dans le champ de l'enquête Revenus fiscaux. Ainsi, l'enquête Emploi en continu compte environ 54 000 logements visités chaque trimestre, représentant environ 35 000 ménages « ordinaires », soit près de 75 000 personnes de 15 ans ou plus enquêtés. Tous les individus de 15 ans ou plus y résidant habituellement sont alors interrogés de façon très détaillée sur leur situation professionnelle. Chaque individu interrogé représente environ 700 personnes compte tenu de la taille d'échantillon. Une des personnes présentes dans le logement au moment de la visite de l'enquêteur peut répondre pour les autres.

L'appariement des données

Il a pour objectif de retrouver les déclarations fiscales remplies par des personnes ayant répondu à l'enquête Emploi. La direction générale des Impôts (DGI) livre trois fichiers à l'Insee :

1. Le premier est le fichier d'imposition des personnes ; c'est la liste des contribuables avec leur adresse au 1er janvier N+1.
2. Le deuxième fichier contient l'ensemble des éléments de taxation du foyer fiscal à l'impôt sur le revenu l'année N ; il s'agit de l'ensemble des éléments de la déclaration fiscale n°2042 remplie en mars N+1.
3. Enfin le dernier fichier contient les données relatives à la taxe d'habitation au 1er janvier N+1.

Ces données issues des fichiers administratifs vont, après appariement, être fusionnées avec celles des individus de l'échantillon de l'enquête Emploi en continu collectées tout au long du 4ème trimestre de l'année N.

L'appariement est effectué au niveau individuel sur des données anonymes (c'est-à-dire sans utilisation du patronyme). Les critères utilisés sont le prénom, le mois, l'année et le département de naissance, le numéro de la voie, deux mots directeurs extraits du libellé de la voie à partir de l'adresse, de la commune et du département. Certaines données d'état civil fiscal des conjoints sont par ailleurs utilisées depuis l'enquête 2003.

L'appariement est validé dès lors qu'une déclaration a été associée à un ménage. Trois cas peuvent se présenter pour un ménage enquêté à l'enquête Emploi :

- Tous les membres du ménage de l'enquête Emploi ont été retrouvés sur une ou plusieurs déclarations fiscales (soit environ 90 % de l'échantillon de l'enquête Emploi au 4ème trimestre). Ils sont donc considérés comme « répondants » dans les enquêtes Revenus fiscaux.
- On retrouve une déclaration fiscale pour au moins un des membres du ménage (environ 4 % des cas). Ils représentent ce que l'on appelle une réponse partielle. Pour ces ménages, les déclarations manquantes sont reconstituées. C'est le cas lorsqu'il y a eu mariage, séparation ou décès. Alors, plusieurs déclarations sont à rédiger pour une même année. Si certaines d'entre elles ne sont pas retrouvées, on les simule.
- Aucune déclaration fiscale n'est retrouvée pour aucun des membres du ménage (environ 6% des cas). Ils sont donc qualifiés de « non-répondants » dans les enquêtes Revenus fiscaux.

Si les ménages non répondants sont très particuliers (c'est-à-dire qu'ils n'ont aucun équivalent dans la population répondante), alors on court le risque de déformer la population de façon significative si l'on ne tient pas compte de ces non-réponses et de la particularité des non-répondants. Les ménages d'étudiants de moins de 26 ans au 31 décembre de l'année N (personnes seules ou couples d'étudiants) posent ce problème. Néanmoins, même quand on ne trouve pas de déclaration (pour les personnes en échec d'appariement), ce qui est le cas de l'immense majorité d'entre eux, on ne les exclut pas. Il s'agit en effet d'enfants rattachés à la déclaration fiscale de leurs parents, vivant hors du domicile des parents et hors logements collectifs ; ils ne sont donc pas présents dans le ménage de l'enquête Emploi.

Ainsi, l'échantillon de l'enquête Revenus fiscaux compte en moyenne selon les années, depuis que l'enquête a pour socle l'enquête Emploi en continu, 35 000 ménages (y compris les ménages dont la personne de référence est étudiante) soit la totalité de l'échantillon de l'EEC au 4ème trimestre de l'année.

La qualité des données collectées

En utilisant des informations d'origine administrative, les enquêtes Revenus fiscaux donnent *a priori* des estimations plus précises de certains revenus que les autres enquêtes auprès des ménages.

En termes de revenu déclaré, la source fiscale est *a priori* de meilleure qualité que les enquêtes réalisées directement auprès des ménages auxquels on demande de déclarer leurs revenus. Il existe *a priori* peu d'erreurs involontaires dans la déclaration fiscale ; les ménages mettent plus de soin à remplir leur déclaration fiscale qu'à répondre aux enquêtes (pas de problème de sous-déclaration par oubli ou de sous-déclaration liée à l'unité de mesure).

Mais en termes de qualité du revenu disponible, l'enquête Revenus fiscaux ne propose qu'une estimation : les prestations familiales, la CSG et la CRDS sont simulées et non observées.

Le redressement

Comme pour toutes les enquêtes réalisées auprès des ménages, le redressement des résultats de l'enquête a deux objectifs principaux : compenser le biais introduit par les non-réponses et réduire autant que possible les fluctuations d'échantillonnage.

Dans le cas où, pour un ménage de l'enquête Emploi, il n'est retrouvé aucun dossier fiscal, il s'agit d'un échec d'appariement qui est donc assimilable à de la non-réponse totale. Afin de compenser le biais lié à la non-réponse totale, on effectue un redressement en deux étapes :

1. la première consiste à mettre en œuvre un modèle de régression logistique de correction de la non-réponse (hors ménages à personne de référence étudiante de moins de 26 ans au 31 décembre de l'année N) ;
2. la seconde consiste en une repondération à l'aide d'une méthode usuelle de « calage sur marges ». Cette méthode consiste à modifier les poids des ménages répondants de telle sorte que le total de certaines variables, en l'occurrence celles citées ci-dessous, estimé à partir des répondants de l'échantillon, soit égal au « vrai » total (ou « marges ») connu par ailleurs (par les [enquêtes annuelles de recensement](#)).

Les marges sont les effectifs issus de la pyramide des âges des individus vivant en « ménages ordinaires » estimée au milieu du 4^{ème} trimestre à partir des résultats des enquêtes annuelles de recensement de la population de deux années consécutives (y compris les ménages dont la personne de référence est étudiante de moins de 26 ans) et les masses des trois types de revenus issues des fichiers fournis par la direction générale des Impôts : les traitements et salaires, les retraites et les revenus d'indépendants.

Par exemple, pour l'enquête Revenus fiscaux 2005, les marges sont les effectifs estimés au milieu du 4^{ème} trimestre à partir des résultats des enquêtes annuelles de recensement de la population 2005 et 2006 ainsi que les masses des trois types de revenus issus des données fiscales de l'année 2005.

Par ailleurs, il est affecté, par imputation, aux ménages d'étudiants de moins de 26 ans, un revenu observé dans la déclaration fiscale d'un individu comparable, en procédant comme suit :

- Parmi les appariements réussis, on identifie des ménages au sein desquels une personne présente dans la (ou les) déclaration(s) fiscale(s), en fait une personne à charge, est absente du ménage de l'enquête Emploi. On suppose que les personnes ainsi identifiées sont des étudiants qui vivent dans un autre ménage (souvent seul).
- En prenant soin de considérer des individus comparables en termes de catégorie socioprofessionnelle et d'âge de la personne de référence (informations présentes dans l'enquête Emploi), on attribue alors aux étudiants en échec d'appariement les revenus des personnes identifiées dans l'étape précédente. Leurs revenus sont donc probablement sous-estimés dans la mesure où l'on ne prend pas en compte les transferts qu'ils peuvent recevoir de leurs parents.

Les estimations des prestations sociales

Cette phase ne concerne que les ménages qui sont dans l'enquête Revenus fiscaux, c'est-à-dire ceux pour lesquels on a retrouvé au moins une déclaration fiscale et les ménages étudiants de moins de 26 ans. Les [prestations sociales](#) étant des revenus non imposables des ménages, ils ne figurent pas dans les fichiers fiscaux ; ils font donc l'objet d'estimation sur barème ou par simulation économétrique.

On peut distinguer trois niveaux d'imputations :

- ✓ Le premier concerne les [allocations familiales](#) de base ou sous condition de ressources : elles dépendent du nombre et de l'âge des enfants. Ces informations étant fournies par l'enquête Emploi, l'estimation se fait par l'application simple du barème en vigueur pour l'année d'enquête. On peut considérer que la qualité de ces estimations est assez bonne.
- ✓ Le deuxième niveau d'estimations recouvre des estimations de prestations dont la perception dépend de paramètres plus complexes, non observés ou observés partiellement (par exemple le handicap individuel, le chômage, l'histoire professionnelle...). C'est le cas notamment des [minima sociaux](#) : [revenu minimum d'insertion](#) (RMI), allocation de parent isolé (API), allocation pour adulte handicapé (AAH), minimum-vieillesse. La simulation se fait en deux temps. Les personnes éligibles sont déterminées en fonction de leur revenus et de leur composition familiale. Si le nombre d'éligibles ainsi estimé est plus important que celui des allocataires recensés par ailleurs dans les données de la caisse d'allocation familiale (CAF), on tire alors au sort des personnes éligibles qui seront exclus de l'exercice de simulation. On affecte aux autres la prestation à laquelle ils ont droit. Cette méthode de simulation est donc un peu grossière et ne garantit pas une qualité parfaite sur des sous-populations particulières.
- ✓ Le troisième niveau concerne les imputations de l'[allocation logement](#). On estime d'abord un montant de loyer à partir de la dernière enquête Logement disponible (*par exemple, pour l'enquête Revenus fiscaux 2005, on a utilisé l'enquête Logement de 2002*). Le montant du loyer affecté aux ménages de l'enquête Revenus fiscaux est estimé économétriquement à partir de variables sociodémographiques et du revenu du ménage. Puis, on applique le barème de l'allocation logement en fonction de ce loyer estimé, du revenu et de la composition du ménage. La simulation dépend donc de modèles économétriques prédictifs et d'observations réelles.