

Sources et méthodes

Enquête Revenus fiscaux et sociaux (ERFS)

L'enquête Revenus fiscaux et sociaux (ERFS) d'une année donnée fournit une évaluation du revenu disponible des ménages pour l'année considérée, enrichie des informations sociodémographiques de l'enquête Emploi de l'Insee.

Conformément aux recommandations du Conseil national de l'information statistique (Cnis) sur les niveaux de vie et les inégalités, elle contient des enrichissements importants et fait suite aux enquêtes Revenus fiscaux (ERF) en vigueur de 1996 à 2005 dans leur périodicité annuelle.

Ainsi, l'ERFS s'appuie, comme précédemment, sur les données de l'enquête Emploi en continu du 4e trimestre de l'année appariées avec les données administratives (déclarations de revenus) émanant de la direction générale des Finances publique (DGFIP). Mais, les données sont désormais appariées avec celles des organismes sociaux afin de disposer des prestations effectivement perçues par les ménages (prestations sociales). Elle évalue également les revenus générés par des produits financiers non soumis à l'impôt sur le revenu.

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 réside dans les différents concepts de revenus qu'elle permet de mesurer. Ces différentes notions sont exposées dans cette rubrique « *Sources et méthodes* » (voir la note : [les concepts de revenus de l'enquête Revenus fiscaux et sociaux](#)).

La nouvelle série de niveaux de vie et de taux de pauvreté désormais produite avec la série d'enquêtes ERFS n'est pas directement comparable avec celle des années passées, provenant d'ERF. En effet, les méthodes d'élaboration de ces deux dispositifs d'enquêtes diffèrent sensiblement.

Présentation générale de l'enquête

L'enquête Revenus fiscaux et sociaux (ERFS) 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 et les prestations sociales, n'est pas collectée par enquête mais par exploitation de sources administratives pour un échantillon représentatif de l'ensemble de la population.

La production de l'enquête Revenus fiscaux et sociaux se déroule en trois 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^e 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.

Les données des individus de l'enquête Emploi en continu qui ont été appariés avec les fichiers fiscaux, sont dans un **deuxième temps**, complétées par les informations provenant des organismes sociaux : les prestations versées à chaque allocataire l'année N par la caisse nationale d'allocations familiales (CNAF) et la caisse nationale de la mutualité sociale agricole (CCMSA), et les droits versés en N+1 par la caisse nationale de l'assurance vieillesse (CNAV).

Dans un **troisième 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 de corrections de la non-réponse et d'un redressement afin d'être représentative au niveau de la France métropolitaine.

Ainsi, l'enquête ERFS 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^e 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, cités universitaires, prisons, etc.) ainsi que les personnes les plus en difficulté, celles vivant en logement précaire (foyers de travailleurs, etc.) et les personnes sans domicile.

L'enquête Revenus fiscaux et sociaux est représentative des ménages métropolitains (approximativement 97 % 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 et sociaux *stricto sensu* est constitué des ménages répondants à l'enquête Emploi et pour lesquels on a retrouvé au moins une déclaration fiscale ou bien un dossier social (dossier dans un des organismes sociaux).

Par conception, l'enquête est tributaire de la législation fiscale qui peut faire varier le revenu déclaré d'une année sur l'autre. Les évolutions et les caractéristiques mêmes de la 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, revenus des entrepreneurs individuels dont l'imposition est basée sur une évaluation forfaitaire, par exemple). L'enquête dépend également des modifications de la législation sociale en vigueur (modifications des caractéristiques des personnes bénéficiaires, etc.).

L'enquête Revenus fiscaux et sociaux de l'année N reconstitue le [revenu disponible](#) au titre des revenus déclarés pour l'année N et ne correspond pas exactement au revenu disponible réellement perçu sur la période :

Par exemple :

- ✓ *Dans l'enquête ERFS de l'année N, le revenu disponible du ménage est égal au revenu déclaré en N, augmenté des revenus financiers non déclarés imputés l'année N et des prestations sociales perçues, le tout diminué de l'impôt sur le revenu de l'année N. Les prestations reçues sont relatives à la législation en vigueur l'année N, calculées par les caisses d'allocation familiale (CAF) en fonction des revenus déclarés en N-1 (ou N-2) ou des revenus perçus au cours du trimestre précédent et de la composition du ménage tout au long de l'année N (avec l'âge des enfants au cours de l'année 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 reçues en N.*

Ce décalage entre ce qui est reconstitué dans les enquêtes ERFS 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.

Mais grâce à l'utilisation des informations d'origine administrative, les enquêtes ERFS donnent des estimations a priori plus précises de certains revenus (salaires, pensions, etc.) que les autres enquêtes auprès des ménages dans lesquelles ces revenus sont renseignés par les personnes interrogées (avec des risques d'omissions que cela comporte).

Les améliorations introduites dans la nouvelle série d'enquêtes ERFS

Conformément aux recommandations du rapport de 2007 du Conseil national de l'information statistique (Cnis) sur les niveaux de vie et les inégalités¹, l'Insee a apporté des améliorations méthodologiques à l'enquête Revenus fiscaux (ERF). Cette dernière est donc remplacée par la nouvelle série d'enquêtes Revenus fiscaux et sociaux (ERFS) mise en place à partir des revenus déclarés en 2006 (pour les revenus de l'année 2005 cependant, l'enquête existe en deux versions : ERF et ERFS). Les améliorations introduites concernent l'intégration des prestations sociales réellement perçues par les bénéficiaires, alors qu'auparavant ces montants étaient estimés à partir

¹ voir le [site du Cnis](#), Rapport N° 103, Niveaux de vie et inégalités sociales - Mars 2007 ; accessible dans les rubriques *Documentation*, puis *Les publications du Cnis* et *Rapports du Cnis*.

des barèmes applicables, et également une prise en compte plus complète des revenus des produits financiers.

Une prise en compte des transferts sociaux reçus par les ménages

Le revenu que les ménages déclarent à l'administration fiscale est complété par les revenus sociaux non imposables : prestations familiales, minima sociaux et prestations logement. Jusqu'en 2005 avec l'enquête ERF, ceux-ci étaient estimés sur barèmes, sur la base des revenus déclarés de l'année, et étaient attribués pour l'année entière. Dans la nouvelle série des enquêtes Revenus fiscaux et sociaux (ERFS), les montants réellement perçus par les ménages sur l'année de revenu considérée sont collectés auprès de la caisse nationale des allocations familiales (CNAF), de la caisse nationale assurance vieillesse (CNAV) et de la caisse centrale de la mutualité sociale agricole (CCMSA).

Par ailleurs, le recours aux fichiers des organismes gestionnaires permet d'intégrer, dans la nouvelle série d'enquêtes, des ménages non retrouvés dans le fichier fiscal et qui étaient auparavant exclus du champ de l'enquête (considérés comme non-répondants) alors qu'ils sont globalement plus pauvres que l'ensemble des ménages (1,2 % de l'échantillon).

Une meilleure couverture des revenus du patrimoine

L'ancienne série de l'enquête Revenus fiscaux qui s'appuyait exclusivement sur la source fiscale appréhendait imparfaitement les revenus du patrimoine, principalement en raison des exonérations fiscales ou des mécanismes de [prélèvement libératoire](#). Cela avait pour effet de minimiser la mesure des inégalités, puisque les revenus du patrimoine sont en général d'autant plus élevés que les revenus d'activité sont importants.

Dans la nouvelle série ERFS, sont ajoutés au revenu des ménages, les revenus générés par différents produits financiers non recensés par la source fiscale et estimés à partir de modèles statistiques. Plus précisément, tous les ménages qui déclarent des revenus d'assurance vie ou de PEA au fisc sont identifiés comme détenteurs. Pour les autres ménages, la détention est imputée ou non d'après les résultats de modèle². Pour cela, un montant des différents actifs financiers détenus par les ménages est imputé à partir de l'enquête Patrimoine réalisée par l'Insee³. Le patrimoine total ainsi imputé reste cependant sous-estimé. Les encours des différents actifs (hors assurance vie) sont alors recalés, produit par produit, sur les données macroéconomiques des comptes nationaux et de la Banque de France (BDF). Les revenus générés par les actifs (hors assurance vie) sont ensuite calculés par application de taux de rendement moyens réactualisés chaque année. Pour l'assurance vie, le recalage se fait sur les revenus et non sur les encours. Faute d'information, le recalage des produits financiers est cependant effectué de manière uniforme : il est indépendant du revenu des ménages.

Si la prise en compte des revenus du patrimoine permet de mieux appréhender la dispersion réelle des revenus, on peut penser que l'application d'un taux de recalage uniforme n'en rend pas complètement compte : la distribution des revenus estimée serait encore moins concentrée qu'elle n'est en réalité. Par ailleurs, les revenus soumis au prélèvement libératoire, hormis les contrats d'assurance vie - obligations, bons, OPCVM - restent mal appréhendés. Au total, les enquêtes ERFS permettent une meilleure estimation des indicateurs d'inégalité qui se rapproche de la situation réelle tout en restant probablement encore un peu en deçà de celle-ci.

² La méthode a été revue en 2007 pour mieux prendre en compte la distribution des durées de contrat.

³ La méthode d'élaboration de l'enquête Patrimoine est accessible dans cette rubrique « Sources et méthodes », à l'emplacement suivant : [Source : enquête Patrimoine 2004](#).

Méthodologie et déroulement de l'enquête

L'échantillon de départ

L'enquête Emploi en continu (EEC), socle de l'enquête Revenus fiscaux et sociaux, 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 tirés au 4^e trimestre de l'année N entrent dans le champ de l'enquête Revenus fiscaux et sociaux puisque c'est l'échantillon qui offre le meilleur taux d'appariement avec les fichiers fiscaux. Ainsi, l'échantillon de départ de l'ERFS est constitué, par exemple, de près de 85 800 individus pour environ 37 100 ménages au 4^e trimestre de l'année 2006. Sont inclus dans cet échantillon les individus appartenant à un ménage de l'EEC même s'ils n'ont pas répondu au questionnaire ainsi que ceux ayant répondu à l'EEC mais n'habitant pas en permanence dans le logement.

Chaque individu interrogé dans l'EEC représente in fine environ 700 personnes compte tenu de la taille d'échantillon.

L'appariement des données avec les fichiers fiscaux

Il a pour objectif de retrouver les déclarations fiscales remplies par les personnes ayant répondu à l'enquête Emploi en continu. La direction générale des Finances publiques (DGFiP) 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 1^{er} janvier N+1.
2. Le deuxième fichier contient l'ensemble des éléments de taxation du foyer fiscal à l'impôt sur le revenu (IR) l'année N ; il s'agit de l'ensemble des éléments des déclarations fiscales n°2042 et n°2042 C (complémentaire) remplies en mai-juin N+ 1 au titre des revenus de l'année N.
3. Enfin le dernier fichier contient les données relatives à la taxe d'habitation (TH) au 1^{er} janvier N+1.

Ces données issues des fichiers fiscaux sont « fusionnées » (après appariement statistique) avec celles des individus de l'échantillon de l'EEC collectée tout au long du 4^e trimestre de l'année N.

L'appariement avec le fichier d'imposition des personnes est dit statistique car il est effectué sur des données anonymes au niveau des individus (c'est-à-dire sans utilisation du patronyme). Les critères utilisés sont la commune et le département, le prénom, le sexe, le jour, le mois et l'année de naissance du déclarant, le département de naissance, le numéro de la voie et deux « mots directeurs » extraits du libellé de la voie (les deux plus grands mots du libellé de la voie figurant dans les adresses⁴). Des informations sur l'état civil du conjoint du déclarant peuvent également être utilisées.

⁴ Ces mots directeurs sont tirés des deux adresses que les fichiers contiennent : l'adresse fiscale de taxation au 1^{er} janvier N+1 et la dernière adresse connue si celle-ci est différente (cas des déménagements).

Ensuite, pour les individus retrouvés, les données portant sur les revenus sont extraites du fichier contenant l'ensemble des éléments de taxation du foyer fiscal. Mais pour certains individus appariés les déclarations ne se retrouvent pas systématiquement. C'est le cas, par exemple, des étudiants redevables d'une taxe d'habitation, mais rattachés à la déclaration de revenus de leurs parents.

Dès lors que pour un ménage de l'enquête Emploi une déclaration fiscale est retrouvée (soit environ 94 % des individus de l'enquête Emploi au 4ème trimestre que l'on s'attend à retrouver), le ménage est considéré comme « répondant » à l'ERFS. Bien qu'il s'agisse d'un ménage répondant à ERFS (présence d'au moins une déclaration fiscale retrouvée pour le ménage), des individus du ménage peuvent ne pas être retrouvés. Ils représentent ce que l'on appelle une non-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. En revanche, si aucune déclaration fiscale n'est retrouvée pour aucun des membres du ménage de l'EEC, ces derniers sont considérés comme « non-répondants » dans les enquêtes ERFS.

Les revenus de personnes qui apparaissent dans les déclarations fiscales (conjoints ou personnes à charge) mais qui ne font pas partie du ménage de l'enquête Emploi ne sont pas pris en compte. Il s'agit le plus souvent d'enfants à charge vivant hors de la résidence familiale pour lesquels aucune condition de vie sous le toit du déclarant fiscal n'est exigée (enfants de moins de 25 ans poursuivant des études, enfants de moins de 21 ans, enfants infirmes et résidant en instituts spécialisés) et d'enfants en résidence alternée absents au moment de l'enquête.

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.

L'appariement avec les fichiers sociaux

Les individus des ménages de l'enquête Emploi en continu sont également appariés avec les données des organismes sociaux : caisse nationale des allocations familiales (CNAF), de la caisse nationale assurance vieillesse (CNAV) et de la caisse centrale de la mutualité sociale agricole (CCMSA).

Les caisses d'allocations familiales gèrent le versement de près de 93 % des prestations prises en compte dans l'ERFS. Aussi, la bonne qualité de l'appariement avec les données de la caisse nationale des allocations familiales est un élément essentiel dans l'enquête Revenus fiscaux et sociaux.

Le déroulement de l'appariement avec le fichier CNAF est le suivant. Les données des individus qui ont été appariés avec les fichiers fiscaux sont complétées par les informations disponibles pour chaque contribuable et son épouse : nom d'usage et/ou nom de naissance (informations cryptées), prénom, date de naissance et adresse fiscale de taxation. Les critères utilisés pour l'appariement des individus âgés de 16 ans⁵ ou plus sont : la date de naissance, le sexe, le prénom (sur 3 caractères), le nom (sur 7 caractères), le département et la commune de résidence. Une liste d'identifiants allocataires présents dans ERFS est ensuite restituée à la CNAF. Au 1^{er} trimestre de l'année N+2, la CNAF retransmet à l'Insee l'ensemble des prestations versées aux allocataires de cette liste durant l'année N. Une faible proportion d'allocataires ne figure plus dans le fichier à l'issue de cette

⁵ Âge minimum des allocataires observé dans le fichier de la CNAF.

procédure : des allocataires ont été radiés en cours de l'année et n'ont donc pas bénéficié de prestations au titre de cette année.

La mutualité sociale agricole (MSA) est l'organisme d'assurance sociale obligatoire des salariés et exploitants du secteur agricole. Elle verse ainsi les prestations familiales, au même titre que la CAF, mais aussi l'ASPA - allocation de solidarité pour les personnes âgées – ou le minimum vieillesse⁶ pour la population agricole. Le rapprochement avec le fichier des allocataires de la MSA est réalisé à partir du fichier des allocataires de l'année N fourni à l'Insee. Les individus qui font l'objet de l'appariement sont les individus de l'EEC âgés de 18 ans ou plus⁷, appartenant à un ménage apparié avec les fichiers fiscaux et qui n'ont pas été appariés (en tant que bénéficiaire ou conjoint) avec le fichier CNAF dans le cadre des prestations familiales. Les critères d'appariement sont la commune de résidence, le sexe et la date de naissance. Le fichier des allocataires de la MSA comprend moins d'informations sur les caractéristiques individuelles des personnes que celui de la CNAF.

La caisse nationale d'assurance vieillesse (CNAV) gère l'ASPA et le minimum vieillesse des individus relevant de la branche générale. L'Insee adresse en avril de l'année N+2 à la CNAV un fichier des individus âgés de 61 ans ou plus (contenant le nom, prénom, date et lieu de naissance) issus de l'appariement entre l'enquête Emploi et les données fiscales. La CNAV se charge ensuite de réaliser l'appariement avec ses données et de retourner à l'Insee le fichier comprenant les droits constatés et versés de l'année N+1 (et non de l'année N⁸).

Le traitement de la non-réponse et le redressement

Comme pour toutes les enquêtes réalisées auprès des ménages, les résultats bruts de l'ERFS sont traités pour répondre à deux objectifs principaux :

1. traiter la non-réponse pour compenser le biais introduit par les non-réponses totales et partielles ;
2. effectuer un redressement afin de réduire autant que possible les fluctuations d'échantillonnage.

Traitement de la non-réponse

Non-réponse totale ou partielle ?

Dans l'enquête ERFS, les revenus d'un individu sont donc tirés des déclarations fiscales qui ont été associées au ménage de l'enquête Emploi dans lequel il se trouve. Lorsque pour un ménage de l'enquête Emploi il n'a été retrouvé ni dossier fiscal ni dossier social, ce ménage est considéré comme une « non-réponse totale » : en effet, on ne dispose ni de ses revenus, ni des prestations reçues. C'est le cas pour un jeune qui n'habite plus chez ses parents, mais qui demeure rattaché au foyer fiscal de ceux-ci. Si le logement du jeune fait partie de l'échantillon de l'enquête Emploi, son dossier fiscal n'est pas retrouvé. Les non-réponses totales, excepté les ménages d'étudiants, font l'objet d'un traitement nécessaire pour compenser le biais lié au fait que, dans la pratique, les non-répondants ne se répartissent pas aléatoirement dans l'échantillon.

En revanche, si un dossier fiscal est retrouvé, mais qu'il ne concerne pas toutes les personnes du ménage, alors ce ménage constitue une « non-réponse partielle » : seules les données d'une partie des personnes de ce ménage sont disponibles. De même si un dossier social (données CNAF) est retrouvé sans qu'aucun dossier fiscal ne soit retrouvé, le ménage considéré est aussi une non-réponse partielle. En cas de mariage, de séparation ou de décès dans l'année, plusieurs déclarations

⁶ Les données sur le minimum vieillesse ou l'ASPA versées par la MSA ne sont récupérées qu'à partir d'ERFS 2007.

⁷ Âge minimum des allocataires observé dans le fichier de la MSA.

⁸ C'est pourquoi la sélection de la population de l'enquête susceptible de percevoir l'ASPA ou le minimum vieillesse est restreinte aux personnes âgées de 61 ans ou plus au lieu de 60 ans ou plus.

sont à rédiger pour une même année. Si certaines d'entre elles ne sont pas retrouvées, le ménage figure parmi les non-réponses partielles : pour ces individus leur revenu est de fait incomplet.

S'agissant des non-réponses partielles, les données des dossiers manquants sont imputées, afin de ne pas perdre une partie de l'information. Ces imputations de données fiscales sont traitées au niveau individuel si le revenu correspondant est individualisé dans la déclaration fiscale. Ceci concerne les salaires, les indemnités de chômage, les pensions de retraite, ainsi que les revenus d'indépendants. Les revenus non individualisables ne sont pas imputés : il s'agit des revenus fonciers, des plus-values et des revenus de capitaux mobiliers portés dans les déclarations fiscales associées au ménage.

Le traitement de la non-réponse partielle : imputation des données manquantes

Les non-réponses partielles concernent donc des individus qui sont dans le champ de l'ERFS mais pour lesquels certaines informations manquent alors que d'autres sont disponibles. Il convient de conserver ces individus en opérant des imputations adaptées à chaque type d'information manquante (revenus, cas des mariages, divorces ou décès, etc.). L'objectif de ces imputations est de remplacer la donnée absente par une donnée « plausible ».

Revenus manquants pour les individus ayant un dossier CNAF

La méthode d'imputation diffère selon le type de prestation reçue par les bénéficiaires car les informations disponibles en matière de revenus dans les fichiers de la CNAF varient selon les cas.

Des imputations de revenus sont réalisées pour les individus (allocataires et conjoints) non appariés avec les fichiers fiscaux mais dont on connaît soit les revenus des 4 trimestres de l'année N (car ils ont perçu le RMI ou l'API sur l'année entière), soit les revenus déclarés de l'année N-1 (ayant servi au calcul d'autres droits pour l'année suivante). En effet, les données fournies par la CNAF contiennent les informations sur les ressources trimestrielles déclarées par les bénéficiaires du RMI ou de l'API : revenus d'activités, revenus de stages de formation, allocations chômage... Lorsque ces informations sont disponibles pour les 4 trimestres de l'année, les revenus trimestriels sont cumulés afin d'obtenir le revenu annuel. Pour les autres individus du ménage (ni allocataire, ni conjoint de l'allocataire) qui ne sont pas appariés avec les fichiers fiscaux, des imputations de revenus sont réalisées selon des modèles économétriques (les individus non retrouvés mais appartenant à un ménage retrouvé dans les fichiers fiscaux servent de base pour ces imputations).

De même des imputations sont réalisées dès lors que les fichiers CNAF contiennent les revenus déclarés par les apporteurs de ressources (salaire, indemnités chômage, pensions...) pris en compte pour le calcul des droits aux prestations familiales. Ces revenus sont ceux de l'année N-1, si l'allocataire bénéficie d'une prestation familiale au deuxième semestre de l'année N. Après imputation, un correctif individuel est réalisé pour passer des revenus de l'année N-1 à ceux de l'année N. La correction est l'évolution moyenne des revenus constatée entre N-1 et N pour les individus bénéficiaires dont les ressources sont connues en N et N-1 (source DGFIP pour l'année N, source CNAF pour l'année N-1). En revanche, les revenus connus dans les fichiers CNAF correspondent à l'année N-2 si l'allocataire bénéficie de prestations seulement pour le premier semestre N. Le ménage est alors considéré comme une non-réponse totale car on ne dispose pas d'information fiable qui permettrait d'imputer un revenu.

Individus dont la situation familiale a évolué au cours de l'année

Des individus ont vu leur situation familiale évoluer au cours de l'année de revenu (année N) parce qu'ils se sont mariés, qu'ils ont divorcé ou parce que leur conjoint est décédé. Ces personnes remplissent alors une déclaration fiscale pour chaque période que l'événement délimite. Les périodes de vie commune se traduisent par le dépôt d'une seule déclaration pour le couple. À l'issue de l'appariement des individus de l'enquête Emploi avec les données fiscales, la récupération de l'ensemble des déclarations n'est pas assurée. Il faut alors compléter le revenu de ces individus pour

estimer leur revenu annuel. Cela couvre les catégories suivantes : salaires, chômage, revenus agricoles, revenus industriels et commerciaux, revenus non commerciaux. La méthode retenue consiste à appliquer à la période manquante le rendement observé (revenus déclarés/nombre de jours) sur la période connue. Cette méthode fait l'hypothèse que ces individus ont une situation au regard de l'activité assez stable sur l'année de revenu étudiée. Pour limiter les risques de surestimation du revenu annuel de ces individus, ces valeurs sont majorées par le revenu maximum observé chez les individus de même catégorie socioprofessionnelle qui ont un revenu annuel complet. Cette méthode n'est appliquée que pour certaines personnes : celles qui disposent d'une période observée de revenus supérieure à 31 jours dans la déclaration retrouvée (la période est suffisamment longue pour que la méthode conduise à une imputation fiable).

Imputation des revenus d'activité ou de remplacement

Pour certaines catégories de personnes, il convient, faute d'information retrouvée, de procéder à des imputations de leurs revenus d'activité (salaires et revenus d'indépendants) ou de remplacement (allocations chômage, retraites). Préalablement aux imputations de ces revenus individuels manquants, revenus bien évidemment liés au statut d'activité des personnes, il faut reconstituer précisément l'activité principale des personnes concernées sur l'année entière (individus de l'enquête emploi du 4^e trimestre de l'année N). Aussi, pour chaque trimestre et donc sur l'ensemble de l'année une activité principalement exercée est reconstituée en mobilisant notamment les concepts d'activité de l'EEC : activité au sens du BIT pour la semaine de référence de l'enquête et calendrier rétrospectif (activité déclarée spontanément sur les 11 derniers mois).

La population des personnes à imputer est constituée des individus d'une tranche d'âge donnée et ayant un statut spécifique au regard de l'activité professionnelle, et selon les cas :

- qui ne sont pas retrouvés dans les fichiers fiscaux et ne sont pas non plus allocataires à la CNAF ;

ou bien

- qui ont connu un changement de leur situation familiale (mariage, divorce, décès du conjoint) l'année N et pour lesquels la période observée de revenus dans la déclaration fiscale retrouvée est inférieure ou égale à 30 jours.

Plus précisément, pour les salaires, les personnes dont le revenu est à imputer sont celles âgées de 18 à 65 ans⁹, ayant pendant au moins un trimestre une activité principale de salariés en [CDI](#) ou en [CDD](#). Pour le chômage, la tranche d'âge est la même et les personnes ont au moins un trimestre en situation principale d'inactivité ou de chômage, à condition qu'ils aient déclaré percevoir une allocation chômage dans l'enquête Emploi. Concernant les retraites, la population à imputer est constituée des individus de plus de 55 ans, ayant au moins un trimestre d'inactivité.

Pour pouvoir imputer les revenus manquants de ces personnes grâce à un modèle économétrique, la population de référence (ou population des répondants) est celle dont l'information en matière de revenus est utilisée pour imputer les données manquantes. Elle est constituée des individus ayant les mêmes caractéristiques en termes d'âge et de situation professionnelle, individus quant à eux retrouvés dans les fichiers fiscaux dans lesquels ils ont déclaré, selon le cas : des salaires, des montants d'allocations de chômage ou de préretraite, des pensions de retraite ou d'invalidité. Un certain nombre de variables auxiliaires sont prises en compte dans les modèles économétriques : sexe, diplôme, âge, catégorie socioprofessionnelle, etc.

Des méthodes d'imputation similaires sont mises en œuvre pour imputer les revenus des indépendants, en traitant séparément les revenus agricoles, les revenus industriels et commerciaux et

⁹ On ne peut imputer de revenus aux personnes en emploi âgées de moins de 18 ans car elles sont peu nombreuses et leurs revenus atypiques : à défaut on considère qu'elles n'ont aucun revenu. Pour la même raison, on ne tient pas compte des très jeunes travailleurs dans le modèle d'imputation des revenus.

les revenus non commerciaux. Au sein des revenus agricoles, les exploitants agricoles qui sont au forfait¹⁰ sont traités séparément : leur particularité est qu'au moment de l'appariement avec les fichiers fiscaux, le montant du forfait n'est pas fixé par l'administration fiscale. Les personnes se déclarant « aide familial 11 » ne sont pas intégrées à ces traitements.

Imputation des revenus des ménages d'étudiants

Les ménages d'étudiants de moins de 26 ans (ménages dont la personne de référence est étudiante) font partie de l'échantillon ERFS même s'ils ne sont pas appariés avec les fichiers fiscaux et sociaux. L'hypothèse est que ces étudiants sont rattachés fiscalement à la déclaration de revenus de leurs parents.

On affecte par imputation à ces ménages 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 corrections de la non-réponse totale

Dans le cas où, pour un ménage de l'enquête Emploi, il n'est retrouvé ni dossier fiscal ni dossier social, il s'agit d'un échec d'appariement qui est donc assimilable à de la non-réponse totale. Les résultats de l'enquête ERFS seraient biaisés si l'on ne corrigeait pas la non-réponse, sauf dans l'hypothèse où les non-répondants ont un comportement identique à celui des répondants. Mais, les échecs d'appariement avec les données fiscales et sociales ne se répartissent *a priori* pas au hasard parmi les ménages de l'EEC. Aussi, afin de compenser le biais lié à la non-réponse totale, on effectue les traitements suivants.

La correction de la non-réponse totale consiste à mettre en œuvre un modèle de régression logistique : il s'agit dans un premier temps d'analyser la non-réponse (hors ménages à personne de référence étudiante de moins de 26 ans au 31 décembre de l'année N) pour mieux la corriger. On postule un modèle de comportement des individus face à la non-réponse en modélisant la probabilité d'appartenir à l'échantillon apparié (i.e. d'être répondant) plutôt qu'à l'échantillon non-apparié (i.e. d'être non-répondant) en fonction de caractéristiques des individus et de leur ménage (par exemple : l'âge, la localisation géographique, la catégorie socioprofessionnelle...). Corriger de la non-réponse consiste à effectuer une modification des poids initiaux¹² des ménages pour prendre en compte le fait que tous ne répondent pas. La correction est d'autant meilleure que les poids initiaux sont transformés de telle sorte que les ménages défailants sont représentés par les ménages possédant autant que possible les mêmes caractéristiques.

¹⁰ Le régime du forfait est appliqué dès lors que les recettes TTC de l'exploitation agricole ne dépassent pas un seuil fixé par l'administration fiscale.

¹¹ Les aides familiaux participent aux travaux d'une entreprise familiale (exploitation agricole, commerce, artisan, etc.) sans être ni "indépendants ou employeurs" ni "salariés". Il s'agit essentiellement d'épouses et d'enfants des chefs d'entreprise individuelle travaillant pour leur propre compte.

¹² Ces poids initiaux sont ceux qu'ont les ménages dans l'enquête Emploi.

Le redressement pour fluctuations d'échantillonnage

Lorsque l'on tire au sort un échantillon pour une enquête, il arrive fréquemment que l'échantillon obtenu ne soit pas exactement représentatif de la population de référence, au regard de certains critères importants pour le sujet de l'enquête. Ces fluctuations d'échantillonnage peuvent être redressées par une modification des poids des individus de l'échantillon. Corriger les fluctuations d'échantillonnage suppose d'avoir des informations issues d'une autre source que l'enquête et de meilleure qualité ou mieux encore la « vraie » valeur. Cette repondération est réalisée à l'aide d'une méthode usuelle de « calage sur marges ». Cette méthode consiste à modifier les poids initiaux¹³ des ménages répondants (i.e. les ménages appariés dans le cas d'ERFS) 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](#), l'[enquête Logement](#) et par les fichiers fiscaux).

L'enquête ERFS est ensuite calée sur des données démographiques (pyramide des âges et nombre de ménages) et sur les masses fiscales déclarées. Concernant la pyramide des âges, les marges sont les effectifs issus de la structure par sexe et âge quinquennal de la population des ménages dits ordinaires de France métropolitaine au 15 novembre de l'année N-14. Pour l'enquête sur les revenus 2007, les pyramides des âges utilisées ont été calculées à partir des résultats du recensement de 2006, obtenus par le cumul des enquêtes annuelles de recensement de 2004 à 2008.

Quant au nombre de ménages au 15 novembre N, il est estimé à partir du nombre connu de résidences principales au 1^{er} janvier N et au 1^{er} janvier N+1 (source Insee, enquête Logement). On fait l'hypothèse d'une évolution linéaire du nombre de résidences principales dans l'année. On estime donc un nombre de ménages par interpolation linéaire avec la formule suivante pour l'année N :

$$\text{Nb de ménages au 15/11/N} = [\text{Nb au 1/1/N}] + [(10,5 / 12) \times (\text{Nb au 1/1/N+1} - \text{Nb au 1/1/N})]$$

Le calage sur les masses de revenus issues des fichiers fournis par la direction générale des Finances publiques (fichiers fiscaux pour la France métropolitaine) est réalisé pour les quatre agrégats : 1.traitements et salaires du ménage, 2.retraites, pensions et rentes, 3.revenus d'indépendants et 4. revenus déclarés du patrimoine¹⁵. Ainsi, les agrégats de calage sont fondés sur des concepts de revenus identiques à ceux de l'ERFS. Les masses fiscales exhaustives utilisées sont corrigées avant le calage afin de correspondre au même champ en termes de ménages : on exclut donc des masses sur lesquelles le calage est effectué les individus ne vivant pas en ménages dits ordinaires.

Remarque : les enquêtes ERFS 2005 et 2006 ont été rétopolées, intégrant les améliorations méthodologiques apportées à l'occasion de la production de l'enquête sur les revenus 2007, les mises à jour des données démographiques et les corrections de la CS¹⁶ dans l'enquête Emploi.

¹³ À partir de l'enquête emploi de 2007, cette pondération initiale a été calée sur le recensement de 2006 intégrant la série des enquêtes de recensement de 2004 à 2008. Une rétopolation, consistant en une mise à jour de ces pondérations, a également été opérée sur les enquêtes emploi à partir du millésime 2002. Compte tenu de la rupture de série pour l'ERFS en 2005, seules les ERFS 2005 et 2006 ont été modifiées en conséquence (mise à jour des pondérations initiales).

¹⁴ Cette date du 15 novembre correspond au milieu du 4^e trimestre (EEC du 4^e trimestre sur lequel est assise l'enquête).

¹⁵ Dans la précédente série des enquêtes ERF, le calage était réalisé sur les trois premiers agrégats.

¹⁶ En raison d'une erreur de codage de la CS à 2 chiffres (avec impact sur celle à 1 chiffre), les catégories socioprofessionnelles ont été rectifiées dans les enquêtes emploi. Ces rectifications ont été intégrées dans les enquêtes ERFS de 2005 à 2007.

Glossaire

Ménage

Dans les enquêtes sur les revenus fiscaux (et sociaux), il s'agit de l'ensemble des occupants d'une résidence principale, qu'ils aient ou non des liens de parenté. Le ménage peut ne comprendre qu'une seule personne. Ne font pas partie des ménages les personnes vivant dans des habitations mobiles (y compris les marinières et les sans-abri) ou dans des communautés (foyers de travailleurs, maisons de retraite, résidences universitaires, maisons de détention...).

Niveau de vie

Revenu disponible du ménage divisé par le nombre d'unités de consommation (UC).

Les dépenses d'un ménage de plusieurs personnes ne sont pas strictement proportionnelles au nombre de personnes, grâce aux économies d'échelle issues de la mise en commun de certains biens. Aussi, pour comparer les niveaux de vie d'individus vivant dans des ménages de taille ou de composition différente, on utilise une mesure du revenu par équivalent-adulte ou par unité de consommation, à l'aide d'une « échelle d'équivalence ». L'échelle actuellement la plus utilisée (dite de l'OCDE modifiée) consiste à décompter 1 unité de consommation (UC) pour le premier adulte du ménage, puis 0,5 UC pour les autres personnes de 14 ans ou plus, et enfin 0,3 UC pour les enfants de moins de 14 ans.

Pauvreté monétaire

Un individu (ou un ménage) est considéré comme pauvre lorsqu'il vit dans un ménage dont le niveau de vie est inférieur au seuil de pauvreté. L'Insee, comme Eurostat et les autres pays européens, mesure la pauvreté monétaire de manière relative alors que d'autres pays (comme les États-Unis ou l'Australie) ont une approche absolue. Dans l'approche en termes relatifs, le seuil est déterminé par rapport à la distribution des niveaux de vie de l'ensemble de la population. On privilégie en Europe le seuil de 60 % du niveau de vie médian.

Prélèvement libératoire

En principe, quand les revenus des valeurs mobilières sont imposables, ils s'ajoutent aux autres revenus soumis à l'impôt sur le revenu. Toutefois, les produits des placements à revenu fixe (c'est-à-dire les obligations, les titres d'emprunt négociables, les créances, dépôts, cautionnements et comptes courants, les bons du Trésor et assimilés, les bons de caisse, les bons et contrats de capitalisation) peuvent être soumis à un prélèvement libératoire forfaitaire qui est un **prélèvement à la source**. Lorsqu'il est opéré (laissé au choix des personnes), le prélèvement libératoire a pour conséquence de libérer de l'impôt sur le revenu les produits auxquels il s'applique.

Le prélèvement libératoire est qualifié de forfaitaire car son taux ne dépend pas du revenu global du contribuable.

Prélèvement à la source

Le prélèvement (ou retenue) à la source est un mode de recouvrement de l'impôt, consistant à faire prélever son montant par un tiers payeur, le plus souvent l'employeur ou le banquier, au moment du versement au contribuable des revenus sur lesquels porte l'impôt.

Revenu déclaré

Somme des revenus déclarés au fisc avant abattements. Il comprend les revenus d'activité salariée ou indépendante, les indemnités de chômage, les pensions alimentaires, d'invalidité ou de retraite, et les revenus du patrimoine. Ces revenus sont nets de cotisations sociales et de CSG (contribution sociale généralisée) déductible.

La déclaration de revenus est propre au foyer fiscal. Le revenu déclaré d'un ménage correspond donc à l'ensemble des revenus déclarés des foyers fiscaux qui le composent. Dans le cas où un foyer fiscal est éclaté sur plusieurs ménages (par exemple, un enfant à charge vivant hors de la résidence familiale), on ne prend pas en compte les revenus des individus absents du ménage.

Revenu disponible

Il comprend les revenus déclarés au fisc (revenus d'activité, retraites et pensions, indemnités de chômage et certains revenus du patrimoine), les revenus financiers non déclarés et imputés (produits d'assurance-vie, des livrets exonérés, PEA, PEP, CEL, PEL) et les prestations sociales, nets des impôts directs (impôt sur le revenu, taxe d'habitation, contribution sociale généralisée - CSG - et contribution à la réduction de la dette sociale - CRDS). Les salaires perçus en rémunération des heures ou jours supplémentaires et des heures complémentaires effectuées à partir du 1^{er} octobre 2007 et non imposables sont incluses dans le calcul du revenu disponible. La mesure du revenu disponible présentée ici correspond à un concept microéconomique de revenu mesuré au travers de l'enquête Revenus fiscaux et de l'enquête Revenus fiscaux et sociaux qui permettent d'étudier les disparités de revenus et la pauvreté monétaire. En parallèle, l'Insee publie des données macroéconomiques relatives au revenu disponible brut dans la Comptabilité nationale, qui font référence pour les évolutions globales. Elles ne sont pas directement comparables aux données microéconomiques, ni en niveau ni en évolution.

Unités de consommation

Les dépenses d'un ménage de plusieurs personnes ne sont pas strictement proportionnelles au nombre de personnes, grâce aux économies d'échelle issues de la mise en commun de certains biens. Aussi, pour comparer les niveaux de vie d'individus vivant dans des ménages de taille ou de composition différente, on utilise une mesure du revenu par équivalent-adulte ou par unité de consommation, à l'aide d'une « échelle d'équivalence ». L'échelle actuellement la plus utilisée (dite de l'OCDE modifiée) consiste à décompter 1 unité de consommation (UC) pour le premier adulte du ménage, puis 0,5 UC pour les autres personnes de 14 ans ou plus, et enfin 0,3 UC pour les enfants de moins de 14 ans.