

Documentation du FPR de l'EIR 2016

Producteur : DREES\OSOL

Base de données : fichier de production et de recherche de l'EIR 2016.

I/ Documentation et mise à disposition de l'EIR 2016

Les informations générales sur l'EIR 2016 sont disponibles :

- Dans le [guide d'exploitation de l'EIR 2016](#)
- Sur la [page de l'EIR 2016](#) du site de la DREES

L'EIR 2016 est diffusé au **CASD**, dans son intégralité (les identifiants internes - *newnordre* - sont remplacés par des identifiants externes - *noind*). Pour la version **FPR**, il est effectué une série de regroupements de variables afin de pseudonymiser la base.

II/ Structure des fichiers de l'EIR

L'EIR est un échantillon de personnes retraitées et invalides (collecté tous les 4 ans par la DREES), construit par appariement de données administratives issues de la quasi-totalité des régimes de retraite français. Il contient les caractéristiques individuelles des retraités et les éléments permettant de calculer leur pension.

La DREES constitue trois bases pour son utilisation propre :

1. Une base individus*caisses de retraite. Elle comprend l'information sur chaque pension versée à l'individu par chaque régime (base *AVANT*).
2. Une base individus fournissant une information agrégée tous régimes confondus, y compris les droits dérivés, et comprenant en outre des variables socio-démographiques de type âge, etc. (base *INDIV*).
3. Une autre base individus comprenant une information agrégée pour les pensionnés d'un régime de base et des variables sur la carrière (base *DDIR*).

L'EIR est apparié avec le panel tous salariés (calcul du taux de remplacement) et avec les données sur les revenus déclarés à l'administration fiscale, depuis 2012. Les données issues de l'appariement fiscal (repondérées) seront diffusées dans un second temps.

1-L'unité statistique diffusée dans le FPR

Le fichier de production et de recherche (FPR) de l'EIR 2016 a pour unité statistique l'individu et intègre des variables issues des trois bases présentées ci-dessus. Il comprend tous les individus ayant un avantage de droits directs ou de droits dérivés perçus en rente et versé en décembre 2016 (cela correspond au champ de la base *INDIV*). Il est conçu pour représenter les personnes âgées de 35 ans ou plus au 31 décembre 2016. Ce fichier ne comprend pas les prestations d'invalidité qui sont dans un fichier spécifique (qui pourra être diffusé dans un second temps, une fois les traitements statistiques entrepris sur cette base). Cependant afin d'assurer une bonne comparabilité entre les régimes, des conventions ont été adoptées. Les anciens fonctionnaires devenus invalides sont intégrés avec les autres bénéficiaires d'une pension de retraite lorsqu'ils atteignent l'âge d'ouverture des droits à la retraite. Avant cet âge, ils sont considérés comme invalides et donc hors champ de ce FPR. Il ne comprend pas les prestations perçues en versement forfaitaire unique. Ces informations pourraient être diffusées dans un second temps avec les données sur les revenus.

Le fichier comprend des personnes nées en France ou à l'étranger, quel que soit leur pays de résidence (personnes inscrites à l'état civil et donc nées ou décédées en France ainsi que les personnes inscrites dans un régime de sécurité social, c'est-à-dire les étrangers ayant travaillé ou s'étant fait soigner en France). L'échantillon était, jusqu'à sa vague 2012, constitué de personnes nées au mois d'octobre, en sélectionnant généralement une

génération sur deux. Pour la vague 2016, l'échantillon s'est élargi en s'étalant sur toute l'année au lieu du seul mois d'octobre, ce qui permet de mieux estimer l'impact de certaines réformes qui touchent parfois différemment les individus d'une même génération (cf. guide d'exploitation).

2-La pseudonymisation des variables indirectement identifiantes

Afin d'élaborer le FPR, les variables indirectement identifiantes de l'EIR 2016 sont modifiées.

- Variable de résidence. Dans l'EIR 2016, la variable de résidence demandée la plus précise est le département de résidence qui est une variable au final bien remplie puisque le bureau Retraites peut l'obtenir auprès de plusieurs caisses de retraite. On décide de passer à la notion de **région de résidence** pour être plus précis que la zone mais sans aller jusqu'au département qui localise trop finement les retraités. La Corse est regroupée avec la région Paca.
- Génération. On choisit de **regrouper des générations** pour ne pas qu'il y ait de trop petits effectifs dans certaines modalités. Une première version consiste aux groupements des individus avant 1932 inclus. Cela correspond à des individus qui ont 85 ans ou plus. Notons que pour les générations les plus récentes, les générations 1960, 1961 et 1962 sont conservées même si les effectifs sont un peu faibles, car ce sont des générations qui représentent des années charnières pour le départ à la retraite (retraite à 54, 55 et 56 ans).
 - On en déduit également les **âges de naissance regroupés** (mais en âges quinquennaux, pas en `age_an`, car sinon on pourrait en déduire les mois de naissance)
 - Pour les générations regroupées, on retire les pondérations par génération qui n'ont plus de sens.
- Régime principal. On va **regrouper certaines caisses** comme les deux caisses du RSI (qui sont maintenant l'actuel SSI) d'un côté, ainsi que les professions libérales, et les régimes spéciaux. Les autres régimes sont conservés tels quels.

Le croisement le plus fin que l'on peut effectuer avec les variables indirectement identifiantes et les niveaux de regroupements proposés ici est le croisement de la génération avec le sexe, la région de résidence et la caisse de retraite du régime principal de liquidation. On ne prend pas en compte, en plus, le croisement avec le lieu de naissance : cette variable est très peu informative, on distingue uniquement les naissances en France et les naissances à l'étranger.

On obtient 5 618 croisements, dont 1 604 avec moins de 11 individus. Pour les individus dans un regroupement à effectif trop faible, nous décidons de ne pas faire de traitement particulier (pas de mise à blanc de la région de résidence notamment).

3-Les autres variables diffusées

Outre le montant mensuel brut total de l'ensemble des avantages servis pour le retraité (*monttot*), on diffuse les composantes qu'on estime être les plus utiles pour des travaux de recherche et qui ne portent pas sur des montants trop détaillés (composantes les plus fines) : les effectifs de retraités concernés par ces composantes fines sont soit assez peu élevés, soit la composante en question ne représente pas un intérêt particulier d'étude en tant que tel (par exemple le montant de la majoration au titre de l'article L.814-2 qui n'est plus attribuée, bien qu'elle continue à être servie), tandis qu'en regroupant cette composante avec d'autres elle prend tout son sens (par exemple, l'ensemble des autres majorations ou suppléments).

On décide donc de regrouper certaines variables (comme les majorations autres que la majoration pour trois enfants ou plus, M5, ou encore les montants de minimum vieillesse). Mais on ne cherche pas à flouter les montants de revenus pour ce FPR, ils sont juste arrondis à l'unité. Par ailleurs, il a été décidé de ne pas faire de traitements particuliers sur les hauts revenus.

III/ Correspondance entre fichiers disponibles au CASD et FPR

Le fichier Excel associé à cette diffusion permet de visualiser les variables disponibles dans le fichier de production et de recherche de l'EIR 2016, par rapport aux données mises à disposition sur le CASD.