





# Sommaire

Apparier pour ajouter des revenus dans des sources Apparier pour mesurer la distribution des revenus

statistiques



# Sommaire

Apparier pour mesurer la distribution des revenus Apparier pour ajouter des revenus dans des sources statistiques

### Les sources administratives disponibles

4

- Sources fiscales (DGFiP) sur les revenus déclarés et les impôts payés : depuis 1999 (revenus 1996)
  - Fichier d'identification des personnes (FIP)
  - Déclarations de revenus (POTE)
  - Taxe d'habitation (PLFC)
- Sources sociales exhaustives sur les prestations versées :
  - Caisse nationale d'allocations familiales (Cnaf) : depuis 2015 (rev. 2012)
  - Caisse nationale de l'assurance vieillesse (Cnav) : depuis 2015 (rev. 2012)
  - Mutualité sociale agricole : depuis 2008 (rev. 2005)



## Trois sources statistiques sur les revenus

- ERFS : Enquête sur les Revenus Fiscaux et Sociaux
  - = enquête Emploi emploi en continu (EEC)
    - + fichiers fiscaux + fichiers sociaux
    - + revenus financiers imputés
- SRCV : Statistiques sur les Ressources et les Conditions de Vie
  - = enquête ménage en panel + fichiers fiscaux + fichiers sociaux
    - + revenus financiers imputés
- Exhaustif Filosofi : Fichier Localisé Social et Fiscal
  - = fichiers fiscaux + fichiers sociaux
    - + revenus financiers imputés



- Pas d'identifiant pour faire le lien entre les fichiers
- Des données potentiellement identifiantes : nom, prénom, sexe, date et lieu de naissance, adresse de résidence, composition du ménage/foyer
  - Mais toutes ne sont pas disponibles dans tous les fichiers
  - Elles ne sont pas renseignées de la même manière
  - Elles sont sujettes à des erreurs de saisie
- Des unités statistiques différentes
  - Ménage vs foyer fiscal ou social

## Production de l'ERFS : étapes

7

- La normalisation des fichiers
- L'appariement
- L'enrichissement
- La correction de la non-réponse
- L'imputation des revenus absents des fichiers

## Production de l'ERFS : appariement

8

- La normalisation des fichiers
- L'appariement : retrouver les individus de l'EEC dans les fichiers administratifs
  - Une méthode par tours, avec des clés d'appariement de plus en plus lâches
- L'enrichissement
- La correction de la non-réponse
- L'imputation des revenus absents des fichiers

- La normalisation des fichiers
- L'appariement
- L'enrichissement : attribuer aux individus et aux ménages de l'EEC les revenus qu'ils ont perçus
  - Individualiser les revenus
  - Gérer les différences d'unité statistique (ménage vs foyer fiscal)
- La correction de la non-réponse
- L'imputation des revenus absents des fichiers



## Production de l'ERFS : correction de la non-réponse

- La normalisation des fichiers
- L'appariement
- L'enrichissement
- La correction de de la non-réponse due au défaut d'appariement
  - 2,2 % de ménages non appariés
  - Environ 5 % d'adultes non appariés au sein des ménages appariés
  - 9 % de la masse des prestations sociales est imputée pour corriger du défaut d'appariement
- L'imputation des revenus absents des fichiers (revenus financiers)
- Au final, l'appariement proprement dit prend moins d'un tiers de la production de l'ERFS



- Données agrégées sur le site et les publications de l'Insee
- FPR SRCV et ERFS
  - au Centre Quételet pour les chercheurs
  - diffusés par l'Insee aux autres utilisateurs
- Microdonnées SRCV et ERFS plus détaillées au CASD
- Microdonnées Filosofi au CASD



# Sommaire

Apparier pour mesurer la distribution des revenus

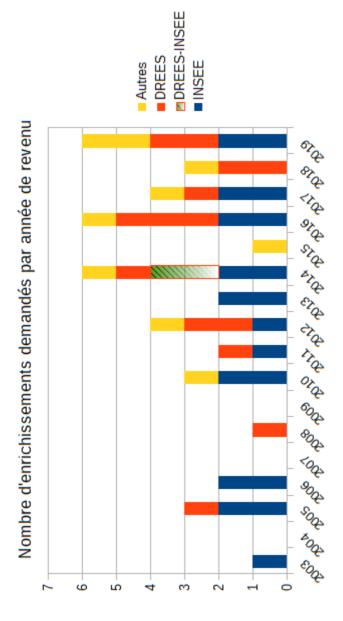
Apparier pour ajouter des revenus dans des sources statistiques



### Des objectifs variés :

- Connaître le revenu dans une enquête généraliste dont il n'est pas la principale variable d'intérêt
  - enquêtes Histoire de vie et patrimoine, Logement, Budget de famille...
- Connaître les revenus d'une population particulière connue via une enquête ou une source administrative
  - les personnes âgées, les bénéficiaires de minima sociaux, les bénéficiaires de contrats aidés, les professionnels de santé, les retraités...
- Une complétude plus ou moins grande (seulement le fiscal/non)
- Un niveau d'agrégation différent selon les enquêtes (individu/ménage/foyer fiscal...)

# ... de plus en plus importants







## Partage des tâches

- Le demandeur normalise (autant que possible) le fichier à apparier
- L'Insee apparie, enrichit
- L'éventuel traitement de la non-réponse est à la charge du demandeur
- Le demandeur est responsable de la diffusion des données



## Cadre juridique

- Les conventions entre l'Insee et les organismes propriétaires des sources administratives prévoient la possibilité d'utiliser ces sources pour enrichir des sources du SSP (et des chercheurs dans le cas des fichiers fiscaux)
- Pour les enquêtes du SSP, l'enrichissement est déclaré dans la fiche de demande d'opportunité et décrit dans le dossier du label
- Le demandeur signe un contrat de sous-traitance avec l'Insee et la DGFIP (la Cnaf le cas échéant)
- Le demandeur est responsable des démarches RGPD
- Il apporte une contribution financière qui correspond au coût des moyens humains engagés pour mener l'enrichissement



Constat : des difficultés pour répondre aux demandes d'enrichissements

### Actions :

- Rationalisation de la chaîne de traitement (2017-2020) et abandon de la méthode d'appariements par tours pour l'utilisation d'une distance calculée sur une clé unique
- Augmentation des moyens (2019)
- Une procédure de planification plus transparente (2020)

## Des questions ?

Julie Labarthe
Cheffe de la division Revenus
et patrimoine des ménages
Julie.labarthe@insee.fr

