
Les doubles comptes dans le recensement en France

Une estimation à partir de l'Echantillon démographique permanent de l'Insee

Laurent Toulemon, Sébastien Durier, Benjamin Marteau
toulemon@ined.fr

Les doubles comptes au recensement

- Omissions et doubles comptes au recensement
- Les données
 - Les données de recensement dans l'EDP
- La méthode
 - des bulletins doubles aux doubles comptes
- Les résultats
 - 1,2% d'habitants comptés deux fois
- La précision de l'estimation
 - Variations aléatoires, erreurs de mesure
- Discussion et conclusion

Introduction

- Un travail dans le cadre du projet Big_Stat
→ http://big_stat.site.ined.fr
- Une communication au congrès de l'UIESP
→ Ici premiers travaux réalisés par Benjamin Marteau



The screenshot shows the homepage of the Big_Stat website. At the top, there is a header with the INED logo (Institut National d'Études Démographiques) on the left, the text "BIG_STAT" in green in the center, and language options "FR | EN" on the right. Below the header is a navigation bar with five tabs: "PROJET", "SOURCES", "RECHERCHE", "LIENS", and a home icon. The main content area is titled "Big_Stat" and contains a paragraph about the project's focus on administrative data. To the right of the main text are two sidebars: "EN SAVOIR PLUS" (More to know) with links to a newsletter, contacts, and a project presentation; and "TÉLÉCHARGEMENT" (Download) with a link to a PDF report submitted to ANR. On the far right, there is a vertical sidebar with icons for search, alphabetical order (A-Z), sharing, and printing.

ined INSTITUT NATIONAL D'ÉTUDES DÉMOGRAPHIQUES

BIG_STAT

FR | EN

PROJET SOURCES RECHERCHE LIENS

Big_Stat

Big_Stat s'intéresse à l'exploitation des données administratives et des fichiers d'enquête mis à disposition de la recherche en démographie. Le projet comprend trois volets.

Recherche. Les données administratives massives issues des données administratives deviennent disponibles pour la recherche. La connaissance sur des situations familiales particulières, difficiles à observer dans les grandes enquêtes de la statistique publique, est renouvelées par le croisement de sources diverses. Tout d'abord, l'attribution des individus à un ménage n'est pas évidente pour les habitants qui vivent de manière habituelle dans plusieurs logements. Il en résulte des doubles comptes et des omissions au recensement, et une définition partielle ou inexacte des situations familiales. Deux situations seront analysées en priorité : les modes de vie précis des enfants de parents séparés, partageant leur résidence entre les domiciles paternel et maternel, ainsi que celui des jeunes adultes revenant de temps en temps chez leurs parents ou partageant leur domicile avec d'autres jeunes dans des « ménages complexes » sans que leur situation familiale ne soit observée correctement au recensement, seront analysées à la lumière des sources disponibles : enquêtes quantitatives et par entretiens, données administratives. L'échantillon démographique permanent de l'Insee, qui regroupe sur un très large échantillon des données d'origine diverses, permet de valider chacune des sources et d'analyser leurs forces et leurs limites.

Mise à disposition. Le deuxième volet du projet consiste à mettre à disposition l'information disponible sur les sources nouvelles. Pour chacune, un site spécifique regroupe une brève présentation de la

EN SAVOIR PLUS

- Pour vous tenir au courant du projet, abonnez-vous à la liste big_stat@listes.ined.fr
- Contacts: Giulia Ferrarì, Laurent Toulemon
- Présentation du projet

TÉLÉCHARGEMENT

- ↓ Projet soumis à l'ANR [PDF | 1,1 Mo.]

Omissions et doubles comptes

- Recensement général de la population
 - On compte tout le monde, une fois et une seule
- Avec le recensement rénové :
 - risque d'omissions moindre
 - risque accru de doubles comptes
- l'EDP pour repérer les doubles comptes
 - Pour les omissions on verra plus tard
- Estimations de population
 - Nombre de ménages, nombre de personnes par ménage

Omissions et doubles comptes

- Recensement général de la population
 - Recensement de 1990 (Coeff 1993) : 0,9% de doubles comptes, 2,7% d'omissions → 1,8% de sous-estimation
 - Aucune enquête post-censitaire depuis
 - Comparaisons avec estimations courantes
 - RP99 : 500 000 de moins que l'estimation courante (Desplanques)
 - Rétablir la cohérence comptable
 - Ne pas changer les données des recensements
 - Introduire un ajustement statistique
- Avec le recensement rénové
 - EAR 2004, RP2006 : 661 000 de plus que prévu
 - Ajustement pour 1999-2006, pas d'ajustement depuis

Estimer les doubles comptes

- Une estimation globale des doubles comptes...
- ... et de la précision de l'estimation
 - Variations aléatoires
 - Erreurs de mesure
- Au-delà du comptage
 - Doubles comptes par sous-groupes (ici par âge)
 - Multi-résidence (TCM) et doubles comptes
- Situation familiale et sociale dans les deux logements
 - Ici situation familiale dans les deux logements

Données : Les recensements dans l'EDP



- Un élargissement de l'échantillon, de 1% à 4%
- Les « Chemises EDP » ne sont plus « nettoyées » des doubles comptes
 - Apparition des bulletins doubles
 - Travaux à l'Insee (Mambetov) : note N° 1253/DG75-F170/D M du 2 juillet 2014 : De 230 à 900 personnes présentes deux fois
 - Bulletins à poids nul, deux bulletins dans le même logement
 - Chaque année, 0,23% des habitants EDP recensés le sont deux fois dans deux logements différents
- Ici l'objectif est d'estimer les doubles comptes
 - Présence simultanée dans deux logements au recensement
 - Pondération spécifique des bulletins doubles

La méthode : des bulletins doubles aux doubles comptes

- Probabilité pour une personne EDP qui a un logement d'être recensé
 - 20% (1/5) dans les petites communes, 8% (2/25) dans les grandes communes (environ, grandes adresses et logements neufs: 20%)
- Probabilité pour une personne EDP qui a deux logements d'être recensé deux fois
 - Même petite commune : $20\% = 1/5$
 - Deux petites communes différentes : $4\% = 1/5 * 1/5 = 1/25$
 - Une petite et une grande commune : $1,6\% = 1/5 * 2/25 = 1/62,5$
 - Deux grandes communes : $0,64\% = 2/25 * 2/25 = 1/156$
 - La même grande commune, deux adresses : $0,64\% = 1/156$
 - La même grande commune, même adresse : $8\% = 2/25 = 1/12,5$

La méthode : des bulletins doubles aux doubles comptes

- En pratique les pondérations observées sont différentes
 - Simples: poids moyen à 7 → petites communes très surreprésentées (70% de l'EDP)
 - Doubles : estimation du poids moyen à 50
 - Coefficient multiplicateur autour de 7 (entre 1 et 12,5)
 - 0,23% → estimation à 1,7% (entre 1,4% et 1,9%, 1,0% et 2,9%)
 - Multi-résidence plus fréquente dans les grandes adresses → poids plus faibles et estimation plus basse
- Sources d'incertitude, biais divers
 - Appariements erronés : deux personnes différentes appariées à la même chemise EDP (jumeaux par exemple) → surestimation
 - Deux bulletins mais un seul apparié à la chemise EDP (négligeable)
 - Double compte moins probable d'une personne si ses deux logements sont recensés la même année (voir plus loin)

La méthode : des bulletins doubles aux doubles comptes

- Incertitude due aux variations aléatoires
 - Avec d la probabilité d'être en double et w les pondérations :
 - $E = E(wd) = \sum_{i \text{ en double}} w(i) = n \bar{w}$
 - $V = V(wd) = V(w)V(d) + V(d)(\bar{w})^2 + V(w)(\bar{d})^2$
 - $V = V(d)(V(w) + (\bar{w})^2)$
 - Si d et w sont indépendants, ce qui n'est pas le cas ici
 - Proportion faible → estimation Poissonnienne corrigée de la variance des poids
 - $E(d) = V(d) \approx n$
 - $V \approx n (V(w) + (\bar{w})^2)$
 - Avec n le nombre de personnes en double et w le poids calculé pour les individus en doubles

Les résultats :

1,2% d'habitants comptés deux fois

- Estimation de la proportion de personnes comptées deux fois
- 2004-2014: 9616 bulletins doubles
 - 2004-2008 : 330 par an ; 2009-2014 : 1330 par an
- Elimination des erreurs
 - Plus de 2 bulletins (32), même logement (427)
 - bulletins à poids nul (2688)
- Reste 6 487 bulletins doubles
 - Ordre de grandeur : 6 500 sur 11 ans
- **Au total : 1,21% des habitants sont comptés deux fois**

La précision de l'estimation

- Variations aléatoires négligeables : +ou-0,046 point
 - Erreur relative : +ou- 4% pour les 11 ans tous ensemble
 - Proportion de doubles comptes: **1,21% IC: [1,16% - 1,25%]**
 - Un peu plus de multi-résidences en grande commune

	Effectifs	Répartition en %	Poids théorique	Poids observé	Rapport obs / th.	Répartition pondérée (%)
Ensemble	2 809 187	100,00		7,07		100,00
Recensés une fois	2 802 700	99,77	7,00	7,01	1,00	98,79
Petite commune	2 045 728	72,82	5,00	5,00	1,00	51,47
Grande commune	756 972	26,95	12,50	12,42	0,99	47,30
Recensés deux fois	6 487	0,23	45,52	36,99	0,81	1,21
Petite, grande	1 903	0,07	62,50	51,53	0,82	0,49
Deux grandes	588	0,02	156,25	116,44	0,75	0,34
Deux petites	2 119	0,08	25,00	25,18	1,01	0,27
Grande, ad. diff	131	0,00	156,25	71,77	0,46	0,05
Même petite	1 441	0,05	5,00	5,01	1,00	0,04
Grande, meme ad.	305	0,01	12,50	11,32	0,91	0,02

Autres incertitudes sur l'estimation

- Les erreurs d'appariement pourraient créer des « faux doubles comptes » (plus qu'en enlever)
- Score d'appariement : moins 0,2 point au maximum
 - Score du plus concordant au moins concordant :
 - Sexe, date de naissance, nationalité, commune de naissance, commune du lieu d'études, études, année d'arrivée, état matrimonial, vie en couple, commune du lieu de travail, diplôme, type d'activité, CS.
 - Score de 5 à 15, seuil à 11 ou plus : on en enlève 17% avec 5 différences ou plus : 816.
 - On enlève aussi 48 jumeaux, dont 45 avec un score haut.
 - Restent 5626 doubles « surs ». **Soit 1,02% en pondéré**
- Les doubles comptes pourraient aussi être plus élevés
 - repérage imparfait des adresses dans les grandes communes (+0,1% ?)
 - Comparaison entre t et $t-1$

Deux années consécutives

- Hypothèse : les doubles comptes sont plus facilement évités quand les deux logements font partie du recensement la même année
- Habitants présents en $t-1$ dans un logement, et présents en t dans un autre logement depuis plus d'un an
 - Pour les années t de 2011 à 2014
 - Impossible dans la même petite commune (0,04%) ou dans la même adresse d'une grande commune (0,02%)
 - Déclarent ne pas avoir changé de logement entre les deux années $t-1$ et t .
 - Pour les autres situations : présomption de déménagement
 - Comparaison à champ comparable; estimation sur 5 vagues

Les sources d'incertitude

- Comparaison $(t-1;t)$, à champ comparable
 - Champ observé : dans un logement en t depuis plus d'1 an et dans un autre logement en $t-1$
 - sauf même petite commune, même adresse dans une grande comm.
 - 1,11% pour les années 2011 à 2014 en $(t;t)$
 - 1,05% dans le champ observé
 - 2,13% pour les années 2011 à 2014 en $(t-1;t)$
 - 2,13% également dans le champ observé
 - **Les doubles comptes sont deux fois plus probable quand les deux logements sont inclus dans le RP dans deux vagues successives**
- Erreur d'appariement: dans 14% des cas on crée un double
 - **fort risque par rapport à 0,23% de bulletins doubles**
- Les variations aléatoires sont négligeables par rapport aux deux sources ci-dessus

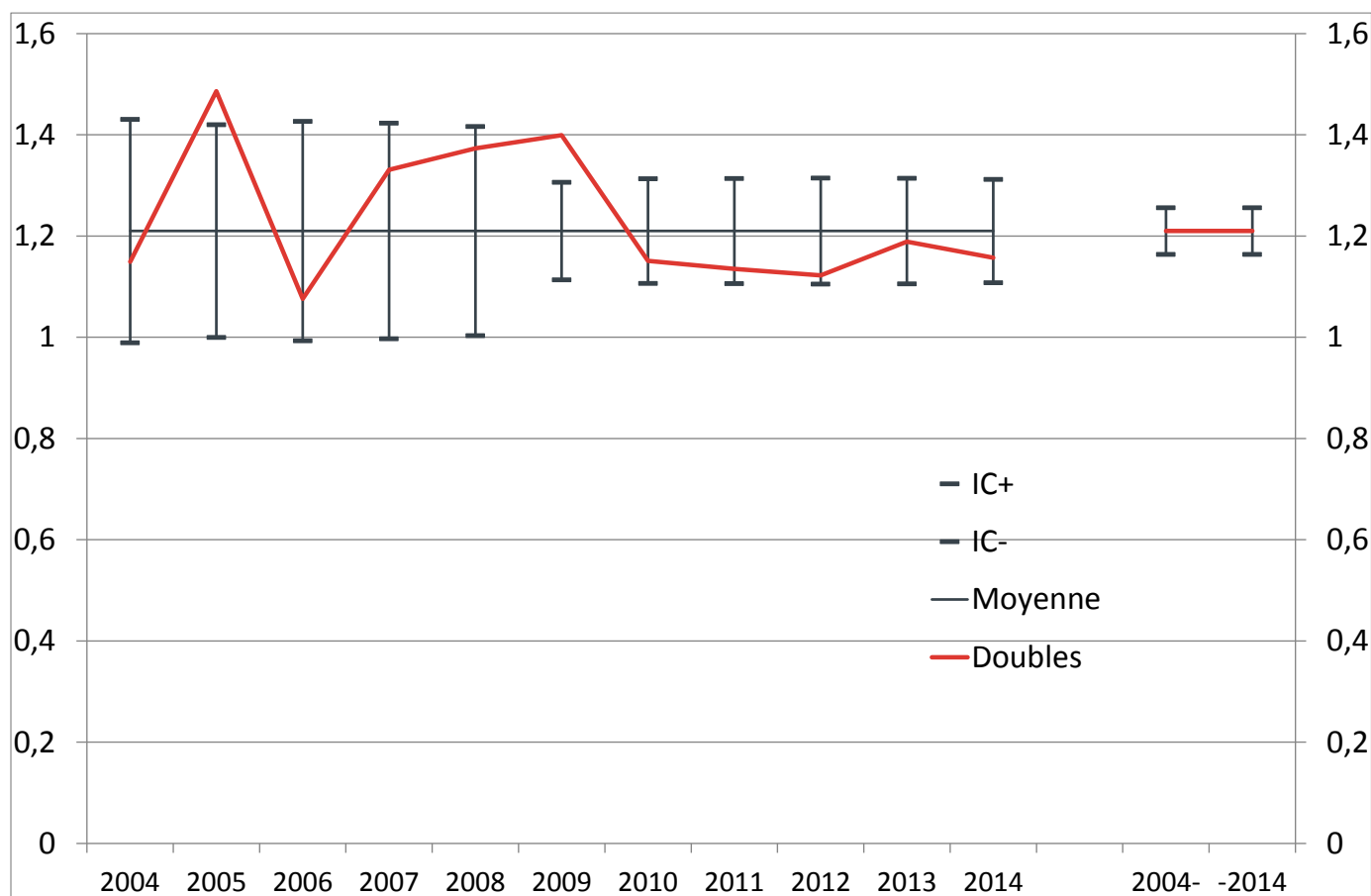
Comparaison avec d'autres estimations

- Hausse modérée par rapport à 1990
 - 1,21% contre 0,7% à 0,9% « probables »
- Plus bas que notre première estimation
 - Un logement dans une « grande adresse »
- Quelques résultats par sous groupe
 - Par année : stabilité des estimations ou légère baisse ?
 - Par âge : très fortes variations
 - Multi-résidences et doubles comptes selon l'âge
 - des profils distincts
 - Situations familiales dans les deux logements
 - Les plus souvent des enfants du ménage (40% des cas)

Peu de changements depuis 2004

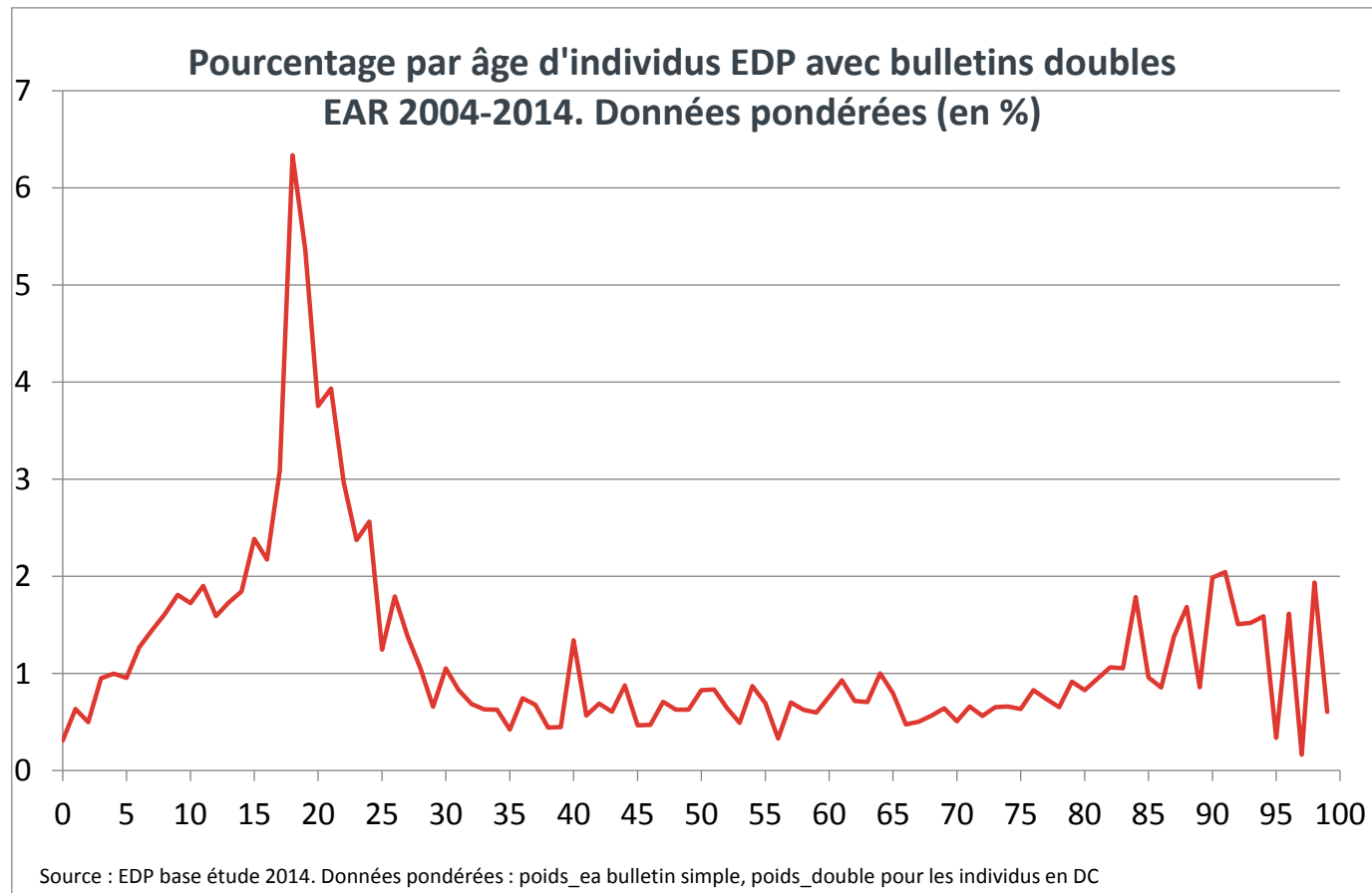
- Baisse après 2009 ?

→ Valider les procédures d'appariement ?



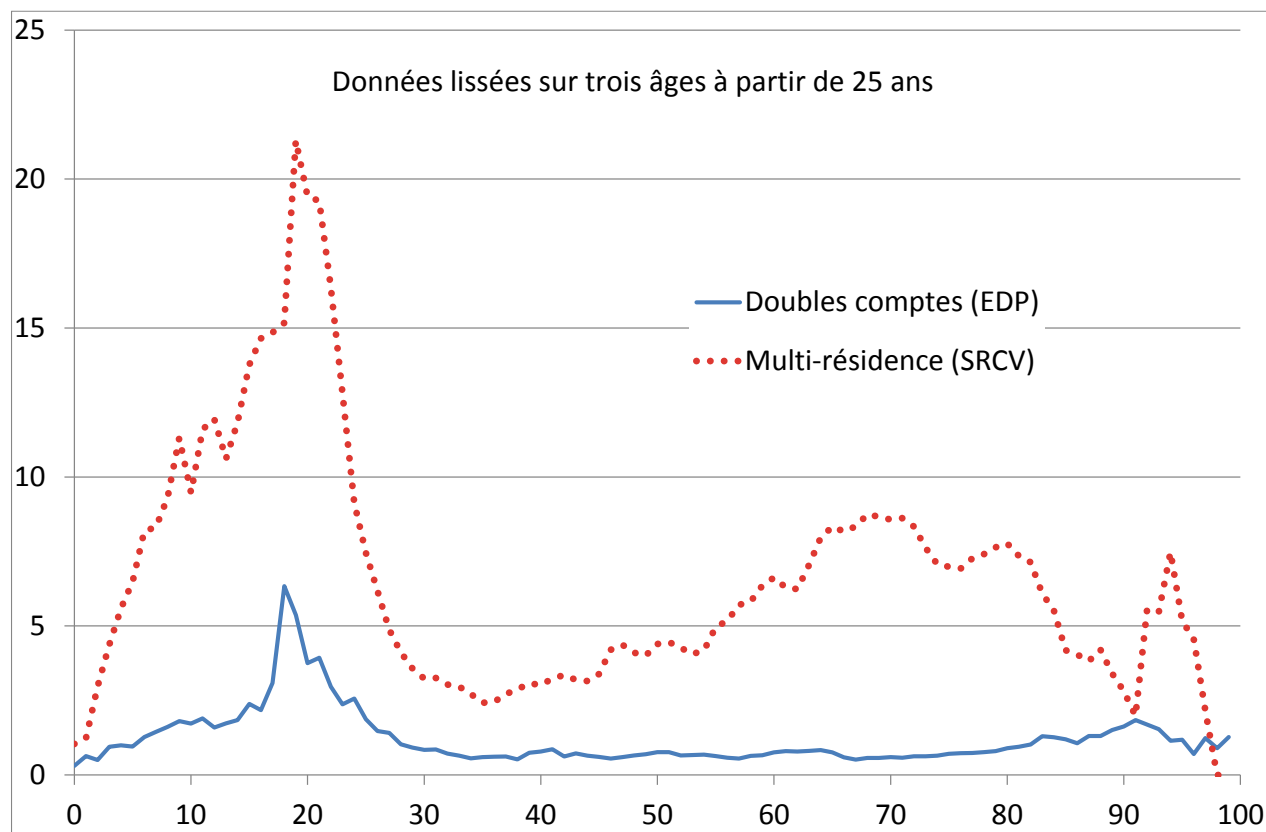
Doubles comptes par âge

- Hausse de 0 à 10 ans, puis pic à 18 ans, faible entre 35 et 80 ans, léger regain aux grands âges



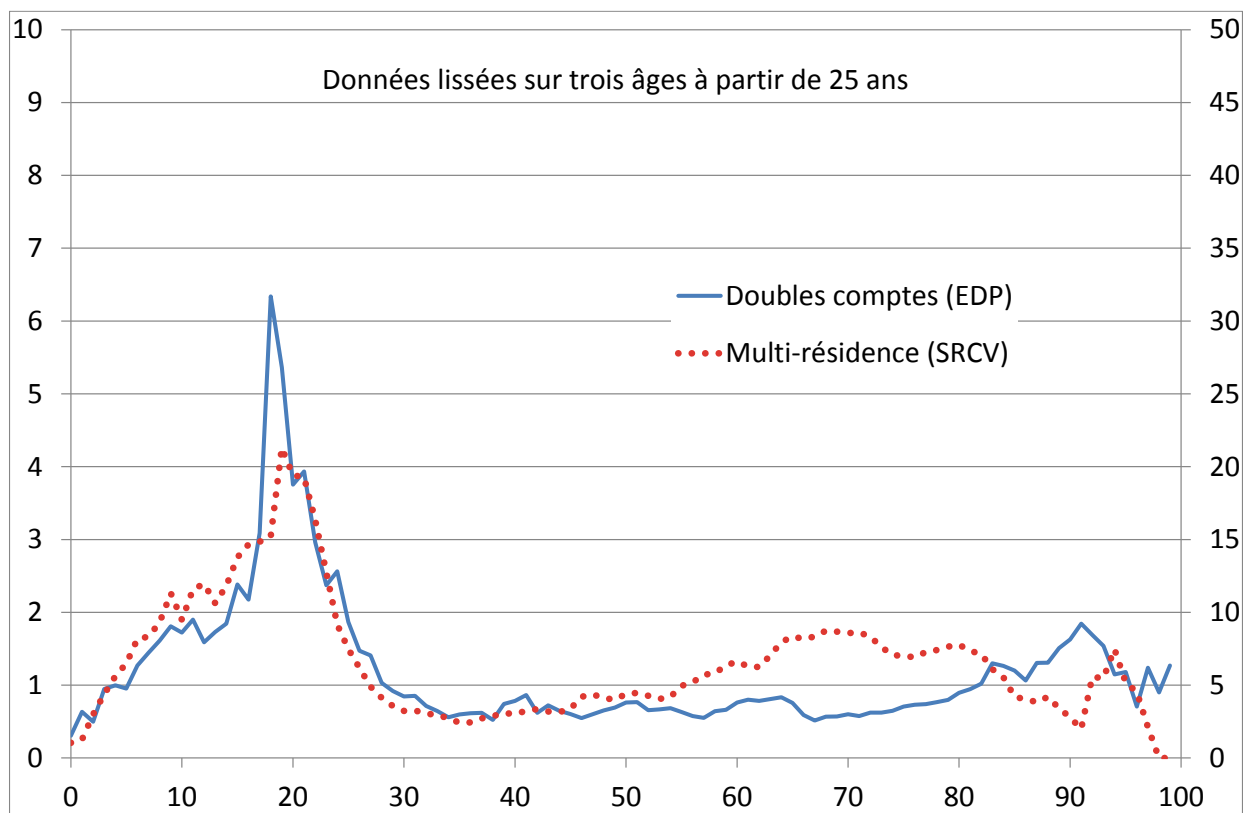
Multi-résidences et doubles comptes

- Des profils similaires : hausse avec l'âge chez les enfants
 - Pas de pic au recensement vers 55-75 ans
 - Décalage pour les jeunes adultes



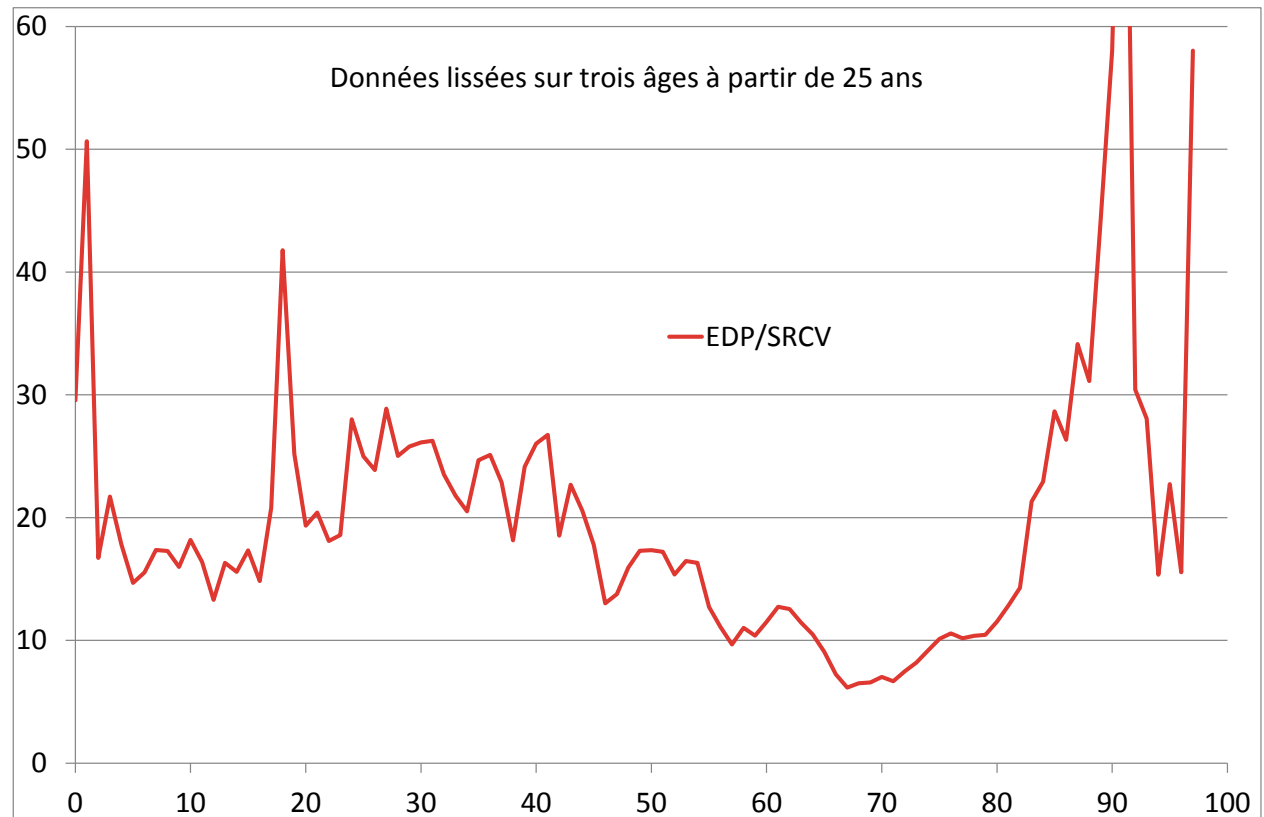
Multi-résidences et doubles comptes

- Axes dans un rapport de 1 à 5 pour superposition des profils
- Des rapports différents pour les enfants, les jeunes adultes, les seniors autour de la retraite, les vieux



Multi-résidence et double compte

- Rapport des doubles comptes aux situations de multi-résidence, selon l'âge (en %). En cas de multi-résidence, les doubles comptes sont les plus fréquents pour...
 - Les très jeunes enfants
 - les jeunes adultes
 - Et les plus vieux
 - Mais pas pour les enfants entre 2 et 16 ans
 - Définition de l'âge à harmoniser



Situation familiale dans les deux logements

- On se restreint aux doubles comptes « surs »
 - Dans 24% des cas la personne est seule dans au moins un de ses logements
- A développer
 - Hommes et femmes
 - Grands groupes d'âge
 - En couple avec le même conjoint...
 - Mêmes autres habitants
 - Situation professionnelle

Situation dans un logement	Situation dans l'autre logement	Répartition (en %)
Enfant	Enfant	25,22
Enfant	Adulte	3,71
Enfant	Isolé	6,09
Enfant	Seul	9,58
Enfant	Hors ménage	15,49
Adulte	Adulte	15,38
Adulte	Isolé	1,72
Adulte	Seul	6,07
Adulte	Hors ménage	5,00
Isolé	Isolé	0,85
Isolé	Seul	1,67
Isolé	Hors ménage	1,99
Seul	Seul	2,99
Seul	Hors ménage	3,60
Hors ménage	Hors ménage	0,65

Prolongements (1/3)

- Des effectifs très suffisants
 - Déjà avec une seule vague: erreur relative < 20%
- Collaboration avec l'Insee pour préciser les estimations
 - Différentes sources d'imprécision et d'incertitude
 - En particulier les questions d'appariement entre recensement et EDP, risque majeur de surestimation des doubles comptes
- Les omissions
 - Les autres sources de l'EDP: sources socio-fiscales
 - Les personnes sont-elles résidentes au sens du recensement ?
- La situation familiale et professionnelle (dans leurs deux logements) des personnes recensées deux fois
 - Les deux familles des enfants de parents séparés
 - Retour sur les situations de multi-résidence

Prolongements (2/3)

- Quel impact sur les estimations de population ?
 - Compliqué, car l'estimation résulte du nombre de personnes par ménage et du nombre de ménages
 - Une personne recensée à tort seule : 1 habitant de plus, mais aussi un ménage de plus → diminution de la population estimée
 - Quand une personne apparaît deux fois, quel bulletin supprimer ?
 - Reproduire les règles de décision
 - Dans le fichier RP, suppression statistique
 - Clones des bulletins supprimés pour supprimer des doubles comptes « non visibles dans l'EDP »
 - Ordre de grandeur : 1,2% d'habitants en moins, entre 0 et 0,3% de résidences principales en moins (ménage d'une personne):
 - **Surestimation de la population entre 0,9% et 1,2%**

Prolongements (3/3)

- Estimation trop haute ? Appariements
 - Score d'appariement: on peut envisager 17% de « faux doubles comptes »
 - Avec 0,3% de faux appariements entre recensement et EDP (estimation à préciser avec des informations hors fichier), on génère 0,04% de bulletins doubles, transformés en 0,2% de faux doubles comptes (18% des doubles comptes seraient des faux)
 - Estimation plausible ?
- Estimation trop basse? Sur cinq vagues
 - 1,21% pour les bulletins doubles la même année ($t;t$)
 - 2,13% pour les bulletins doubles sur deux années ($t-1;t$)
 - Généralisé à ($t-u;t$) pour u de 1 à 4
 - En moyenne 1,9% avec cette correction
- **IC: [1,16% - 1,25%] mais estimation [0,8% - 1,9%]**
- Traiter aussi les omissions
- Nécessité d'un diagnostic partagé ?

Conclusion

- Hausse modérée des doubles comptes (0,7% à 0,9% en 1990, 1,2% avec le recensement rénové)
 - Améliorer le recensement?
 - Exemple des Etats-Unis (2010)
- 7. Does this person sometimes live or stay somewhere else?**

☐ No ☐ Yes — Mark ☒ all that apply.

☐ In college housing	☐ For child custody
☐ In the military	☐ In jail or prison
☐ At a seasonal or second residence	☐ In a nursing home
	☐ For another reason

→ If more people were counted in Question 1 on the front page, continue with Person 6.
- L'EDP permet de valider le recensement
 - Et de compléter les analyses
 - Ainsi que des retours sur l'évaluation de l'EDP
 - **Merci de votre attention**