

La nouvelle planification de l'échantillonnage

Pierre-Arnaud Pendoli
Division Sondages



Plan de la présentation

- Rappel sur le Recensement de la population (RP) en continu
- Description de la base de sondage issue du RP
- Les nouvelles contraintes liées aux NCEE
- Les adaptations nécessaires de l'échantillonnage

Principe du nouveau Recensement (1)

- Recensement en continu => base de sondage rafraîchie d'année en année

- Création de cinq groupes de rotation du RP : chaque année, un seul groupe est impacté par le Recensement

- Distinction entre petites et grandes communes
 - Petites communes :
 - Chaque petite commune est affectée aléatoirement à l'un des cinq groupes de rotation du RP

 - Une année donnée, toutes les communes du groupe de rotation impacté sont recensées exhaustivement

Principe du nouveau Recensement (2)

- Distinction entre petites et grandes communes
 - Grandes communes :
Dans une grande commune donnée, répartition aléatoire des **adresses** entre les cinq groupes de rotation du RP

- Au bout d'un cycle de 5 ans de RP :
 - Recensement exhaustif des logements des petites communes

 - Recensement d'environ 40% des logements des grandes communes

Unités primaires = les zones d'action enquêteur (1)

- Mise en place d'un échantillon maître

- Comment construire les ZAE ?
 - Réaliser une partition du territoire

 - Fraîcheur : chaque ZAE doit contenir des logements des 5 groupes de rotation (GR) du RP

 - Volume : chaque GR doit comporter au moins 300 résidences principales

 - Respect des limites régionales

 - Étendue géographique la plus faible possible

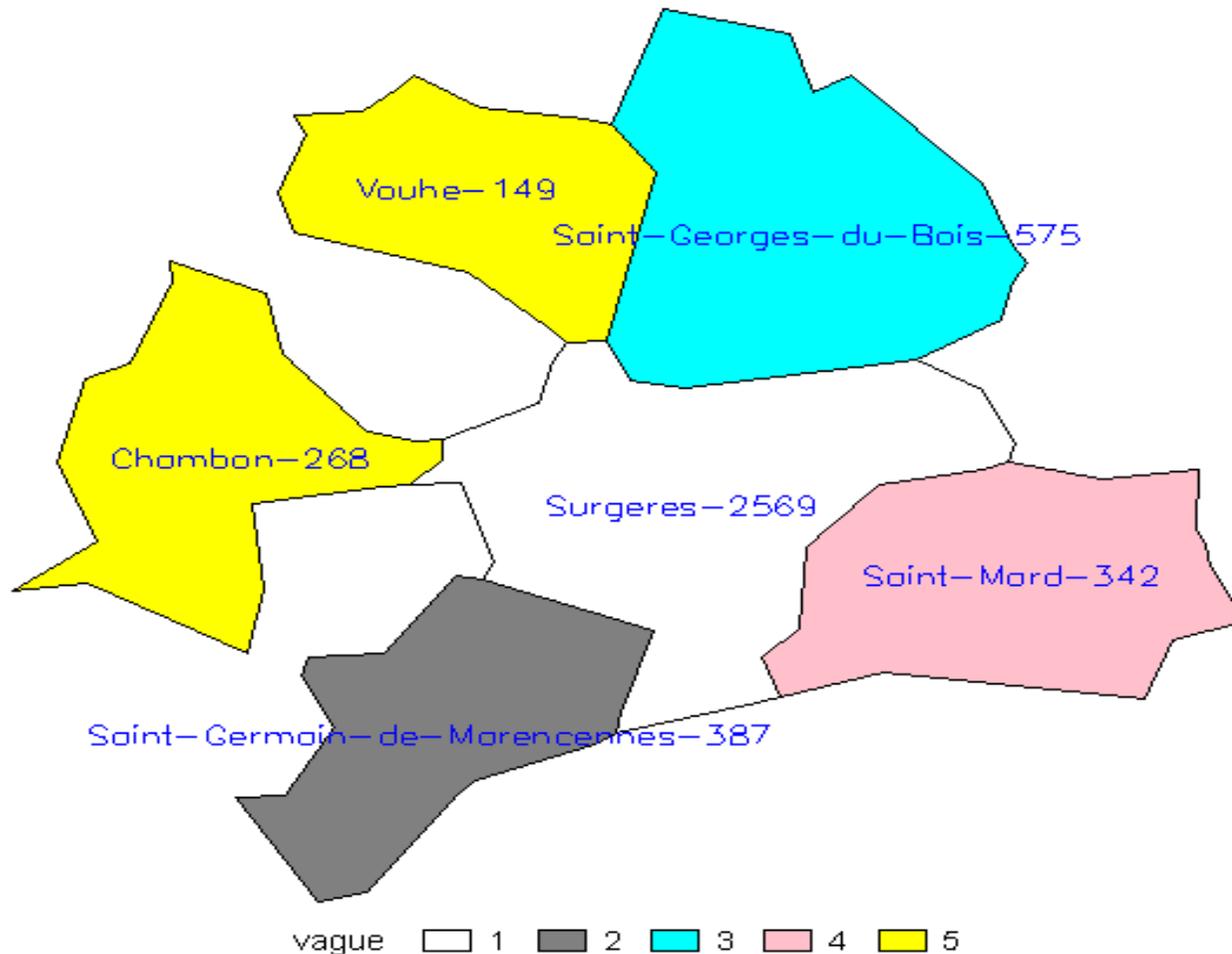
Unités primaires = les ZAE (2)

- Distinction entre petites et grandes communes :
 - En petites communes :
 - Une ZAE est constituée de plusieurs petites communes
 - Une ZAE comporte au moins une commune de chaque GR
 - Algorithme déterministe pour construire les ZAE de manière à minimiser leur étendue
 - En grandes communes :
 - Une grande commune = une ZAE
 - Une ZAE comporte bien des logements de chaque GR

- Résultat : 3 832 ZAE dont :
 - 2 898 ZAE « petites communes » (PC)
 - 934 ZAE « grandes communes » (GC)

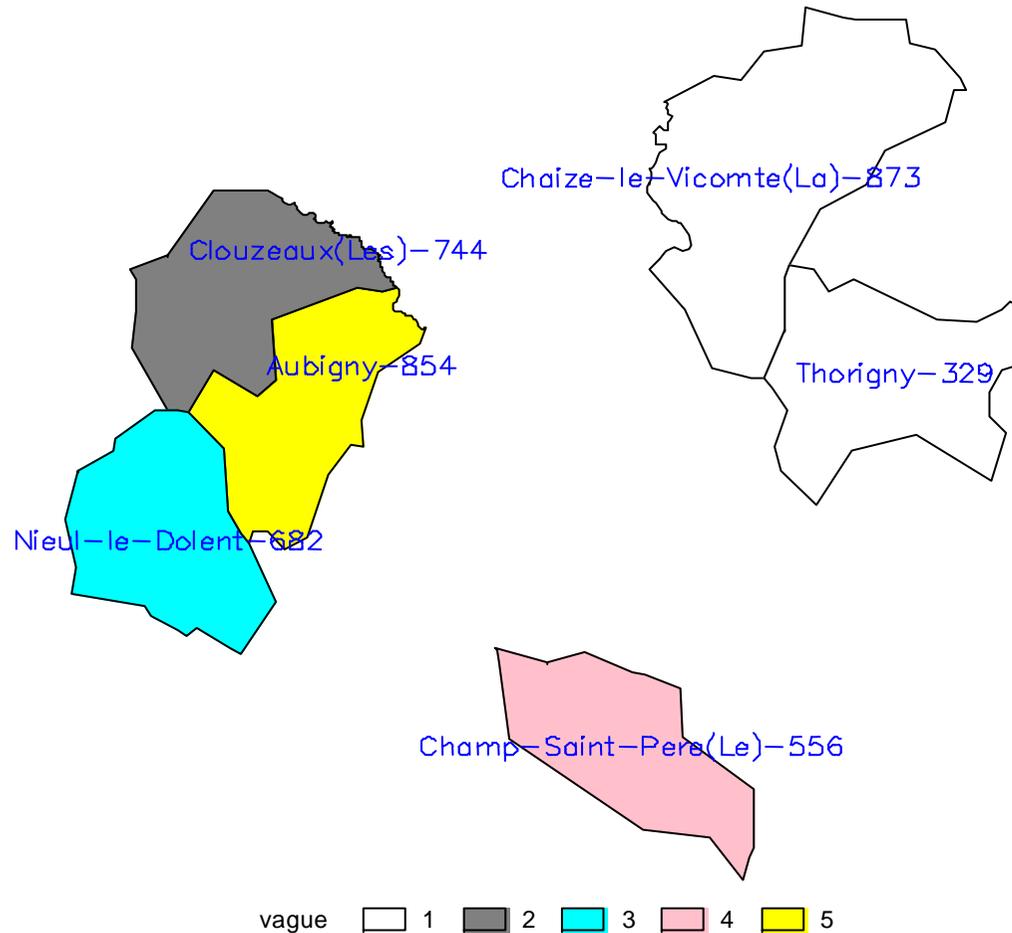
Unités primaires = les ZAE (3)

- Exemple de ZAE PC contiguë Z17434



Unités primaires = les ZAE (4)

- Exemple de ZAE PC non contiguë
Z85046



Tirage de 1er degré (1)

- Objectif du tirage des ZAE (i.e. constitution de l'Echantillon-Maître) : concentrer le tirage des échantillons pour
 1. Diminuer les coûts de collecte
 2. Embaucher un réseau d'enquêteurs fixe

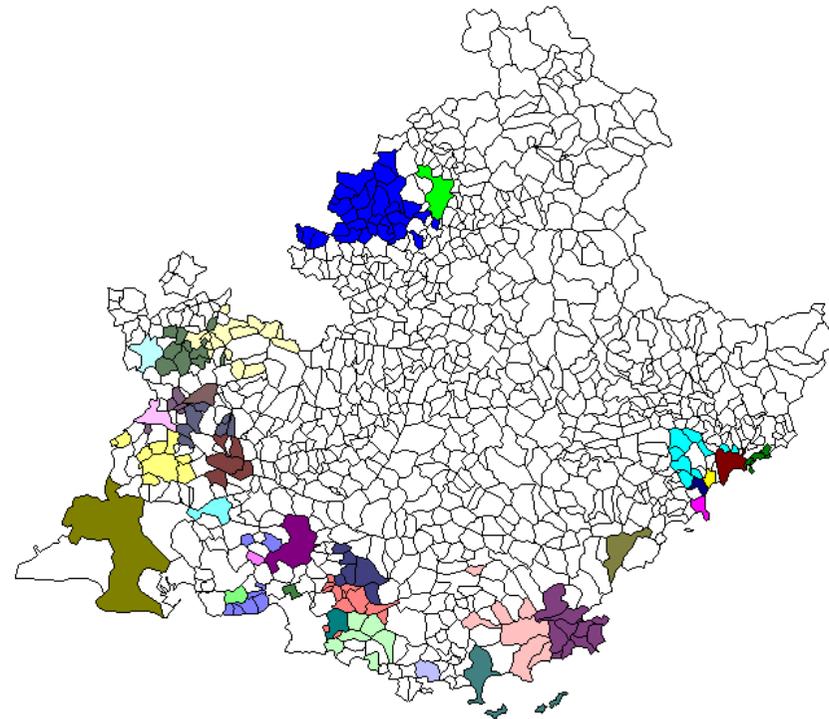
- Méthodologie :
 - Allocation : 567 ZAE
 - Probabilités d'inclusion proportionnelles au nombre de résidences principales d'après le RP 1999
 - Stratification par région avec allocations proportionnelles
 - Equilibrage sur le nombre total de résidences principales, le revenu fiscal total et la taille d'unité urbaine

Tirage de 1er degré (2)

- Résultat : 567 ZAE qui forment l'Echantillon-Maître :
 - 286 ZAE PC
 - 281 ZAE GC dont 80 ZAE dites « exhaustives » (probabilité d'inclusion égale à 1)

- Illustration : ZAE de l'EM en PACA

Répartition géographique des ZAE non uniforme
⇒ problème pour la collecte en cas d'absence d'un enquêteur sur sa zone



Calage des poids de sondage des ZAE de l'EM

- Poids de sondage initial = inverse de la probabilité d'inclusion de la ZAE calculée au moment du tirage des ZAE en 2007, d'après les données du RP 1999
- Les ZAE de l'EM sont représentatives de la population au 01/01/1999
- Chaque campagne de Recensement rafraîchit la base de sondage => nécessité de caler l'échantillon de ZAE EM sur la dernière campagne de Recensement
- En utilisant les poids calés, les ZAE EM sont représentatives de la base de sondage mise à jour

Tirage de 2nd degré : les logements (1)

- Base de sondage annuelle = les résidences principales appartenant :
 1. Aux ZAE de l'EM
 2. Et au groupe de rotation correspondant à la campagne de Recensement la plus récente

- Exception : certaines enquêtes particulières sont tirées dans un cycle complet du RP. Ex. « Cadre de vie et sécurité »

Tirage de 2nd degré : les logements (2)

➤ Méthodologie du tirage des logements

- Tirage à deux degrés : calcul d'allocations par ZAE permettant de minimiser la dispersion des poids de sondage = allocations théoriques

Pour les ZAE PC :

$$W \log_{ZAE}^{PC} = \omega_{ZAE}^{calé} * \frac{N_{ZAE}^{GRi}}{n_{ZAE}}$$

Pour les ZAE GC :

$$W \log_{ZAE}^{GC} = \omega_{ZAE}^{calé} * \frac{\hat{N}_{ZAE}^{GRi}}{n_{ZAE}}$$

Tirage de 2nd degré : les logements (3)

➤ Méthodologie du tirage des logements

- Possibilité de mettre des bornes min et max pour prendre en compte la charge des enquêteurs
- Tirage systématique dans chaque ZAE
- Tirage en deux phases dans le cas d'enquêtes à champs restreints ou à sous/sur représentations (stratification)

Les nouvelles conditions d'emploi des enquêteurs (NCEE)

➤ Le nouveau cadre prévoit de :

1. Connaître la charge d'enquête pour l'ensemble des enquêtes qui seront sur le terrain l'année N dès la fin du 1^{er} semestre de l'année N-1

2. Estimer pour chaque enquêteur la charge de travail induite et vérifier qu'elle est compatible avec sa quotité annuelle préalablement fixée

=> À partir de 2013, les allocations par ZAE pour les enquêtes de l'année N doivent être calculées au 2^{ème} trimestre de l'année N-1

L'impact des NCEE en amont de la collecte (1)

- Pour les enquêtes collectées au 1er semestre 2015, aucun changement par rapport aux années précédentes :
 - Tirage des échantillons aux mois de mai et juin 2014...
 - ... dans la campagne RP 2013 (disponible depuis février 2014)
- En effet, on ne dispose pas d'une base de sondage plus « fraîche » avant le début de la collecte

L'impact des NCEE en amont de la collecte (2)

- Pour les enquêtes collectées au 2nd semestre 2015 :
 - Tirage des échantillons **aux mois de mai et juin 2015...**
 - ... dans la campagne RP 2014 (disponible depuis février 2015) => on utilise la base de tirage la plus « fraîche » possible
- Problème : les allocations de FA par ZAE doivent être connues **au mois de juin 2014** pour les enquêtes sur le terrain en 2015

Solution : anticiper les allocations par ZAE

- Pour une enquête donnée, les allocations par ZAE dépendent :
 - de la taille de l'échantillon
 - de la taille de la ZAE (nombre total de résidences principales ou effectifs dans le champ de l'enquête, d'après la dernière campagne de Recensement)
 - du poids de sondage de la ZAE

- Pour les enquêtes collectées au 2nd semestre de l'année N, il s'agit de s'approcher des allocations théoriques qui seraient tirées dans la campagne RP N-1 :
 - à partir des tailles des ZAE connues dans une ancienne campagne RP
 - et des poids de sondage associés

Quel millésime de campagne RP utiliser ? (1)

- La campagne de Recensement connue la plus récente ?
Idée : anticiper les allocations à tirer dans la campagne N à partir des effectifs présents dans la campagne N-1
 - Avantage : donne la représentation des ZAE la plus récente, i.e. correspondant le mieux à la représentation présente des ZAE dans la population
 - Inconvénient : grandes disparités entre les groupes de rotation du RP au sein d'une même ZAE
 - **Entre deux campagnes successives, le ratio du nombre de logements dépasse les 3 fois pour plus de la moitié des ZAE**
 - **La moitié des ZAE « PC » ont été contraintes au moins une fois à la borne supérieure et au moins une fois à la borne inférieure pour les 4 dernières campagnes RP**

Quel millésime de campagne RP utiliser ? (2)

- La dernière campagne de Recensement qui correspond au même groupe de rotation?

Idée : anticiper les allocations à tirer dans la campagne N à partir des effectifs présents dans la campagne N-5

- Pour les ZAE PC, la campagne N-5 correspond aux mêmes communes que la campagne N

Hypothèses :

- A cinq ans d'intervalle, les tailles relatives des communes ne sont pas bouleversées ...
- et leurs caractéristiques (tranches d'âge, PCS, ...) non plus

Comparaison des solutions envisagées (1)

- Les allocations « anticipées » s'écartent-elles des allocations théoriques issues du tirage ?
- Calcul par ZAE de l'écart relatif entre allocations théoriques et « anticipées »
- Comparaison « toutes choses égales par ailleurs » (**poids de sondage initiaux utilisés dans tous les cas**) entre le calcul d'allocations :
 - via la campagne RP N-1
 - via la campagne RP N-5

Comparaison des solutions envisagées (2)

- Simulation d'un échantillon en 1 phase de 30 000 logements, tirés dans l'Echantillon-Maître :
 - Allocations théoriques du tirage dans la campagne RP N = 2013, avec les poids de sondage initiaux des ZAE
 - Allocations « anticipées » calculées à l'aide des poids de sondage initiaux des ZAE, depuis :
 - la campagne RP N-1 = 2012
 - la campagne RP N-5 = 2008

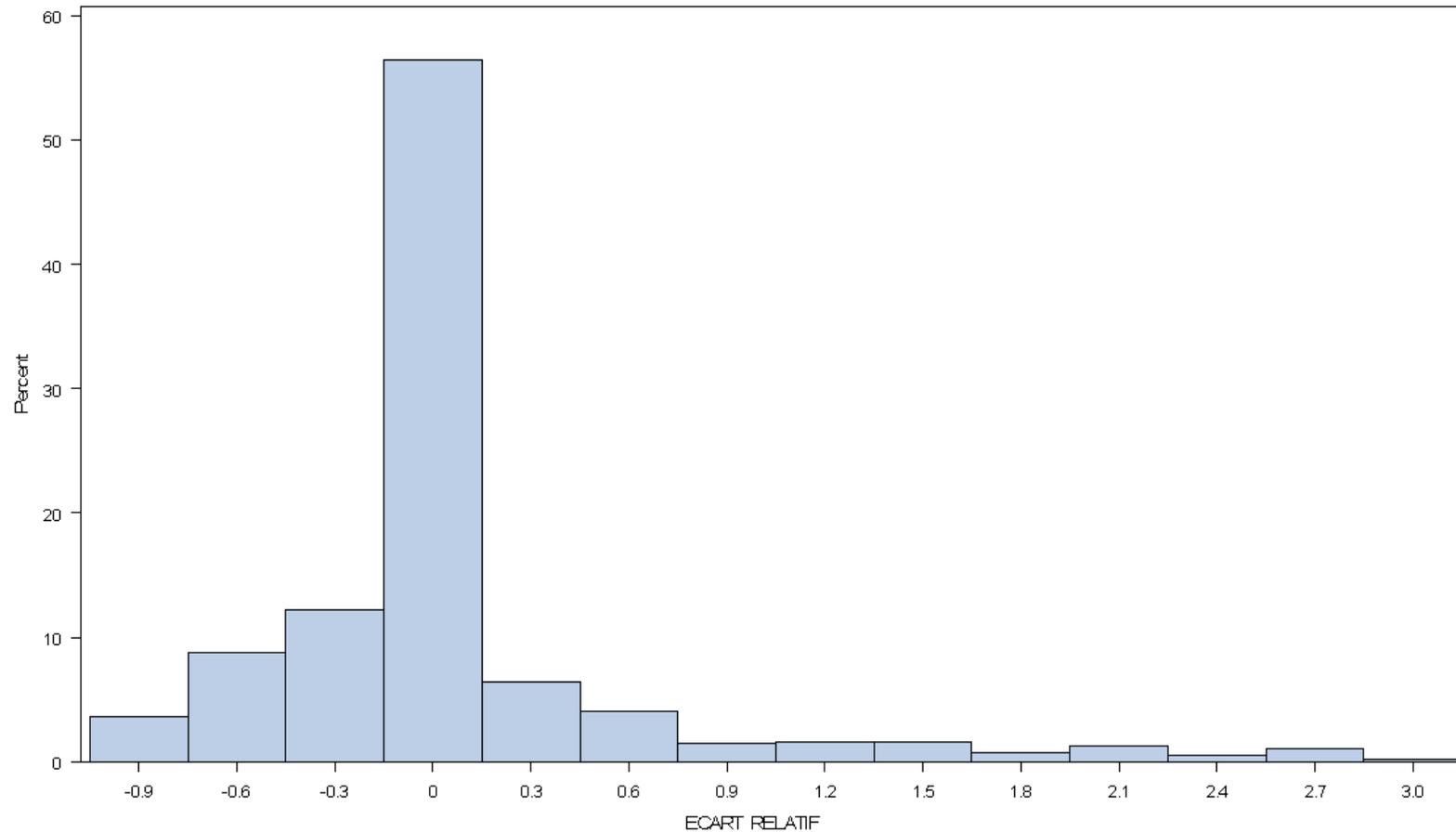
- Ecart relatif calculé :

$$Ecart\ relatif_{ZAE} = \frac{(Allocation\ anticipée_{ZAE}) - (Allocation\ théorique_{ZAE})}{Allocation\ théorique_{ZAE}}$$

Comparaison des solutions envisagées (3)

➤ Résultats

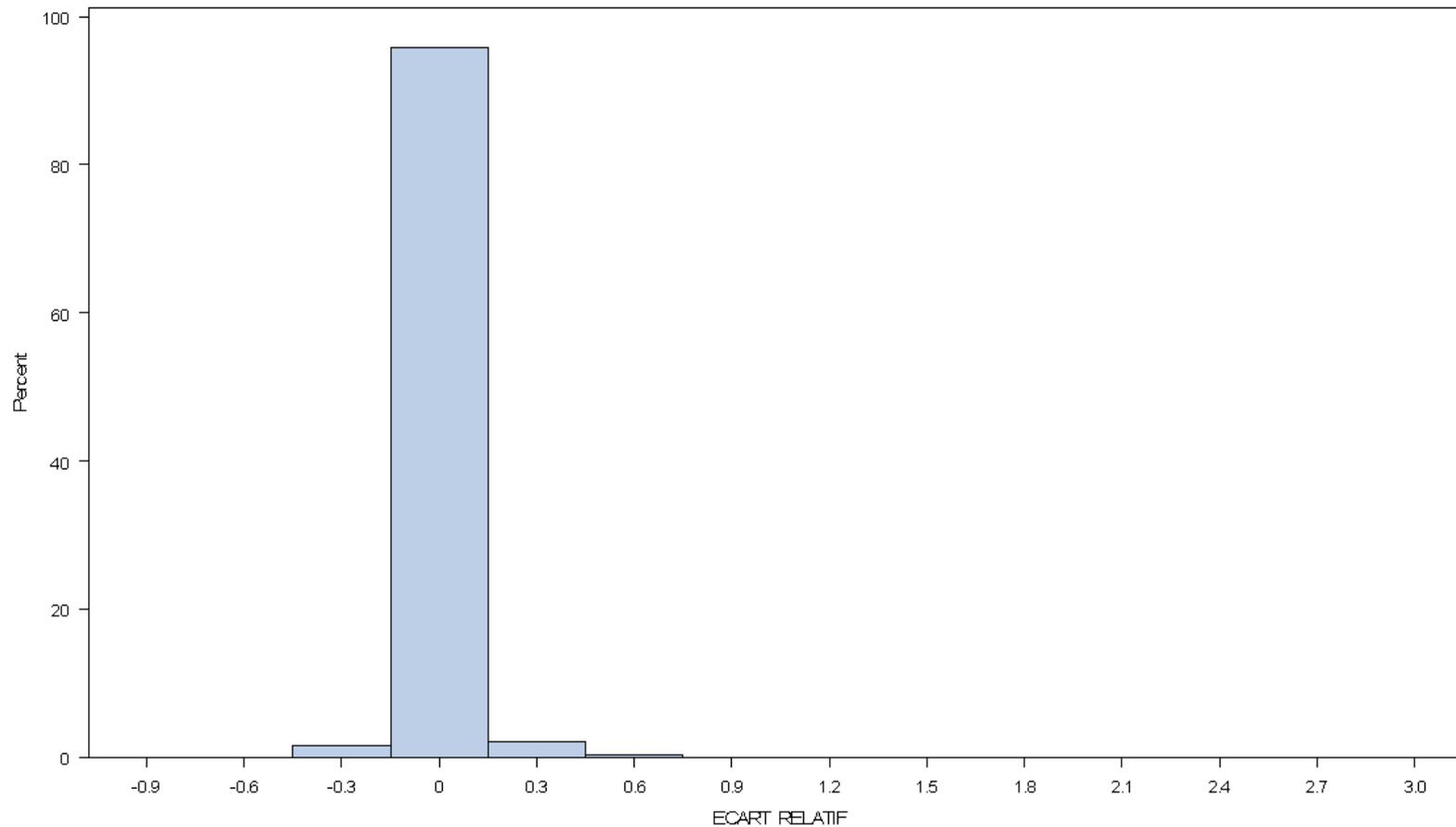
- A partir de la campagne RP N-1 = 2012



Comparaison des solutions envisagées (4)

➤ Résultats

- A partir de la campagne RP N-5 = 2008



Comparaison des solutions envisagées (5)

- A cinq ans d'intervalle, les tailles relatives des ZAE, pour un même groupe de rotation, évoluent peu
- La campagne N-5 permet de bien anticiper la taille des ZAE pour la campagne N
- La campagne N-5 est donc bien meilleure pour estimer les allocations par ZAE

Quel jeu de poids utiliser ?

➤ Printemps 2014 :

- Calcul anticipé des allocations pour les enquêtes du 2nd semestre 2015 (tirage dans la campagne RP 2015)
- La campagne RP 2015 n'est pas encore disponible => les poids des ZAE calés sur la campagne 2015 n'existent pas

➤ Recourir aux poids de sondage initiaux des ZAE ?

➤ Ou recourir aux poids des ZAE calés sur la campagne N-5 = 2010 ?

Comparaison des solutions envisagées (1)

- Quel est l'effet, « toutes choses égales par ailleurs », des pondérations utilisées sur la qualité des allocations ?
- Pour un tirage de 30 000 logements dans la campagne RP 2013, comparaison entre les allocations :
 - calculées à partir des poids calés sur la campagne 2013, i.e. les allocations théoriques
 - calculées à partir de pondérations plus anciennes, i.e. les allocations « anticipées »
 - i. les poids de sondage des ZAE calés sur la campagne RP N-5 = 2008
 - ii. les poids de sondage initiaux des ZAE

Comparaison des solutions envisagées (2)

- Résultats : calculs par ZAE des écarts relatifs entre allocations théoriques et « anticipées »

Quantile	Poids de sondage initiaux	Poids de sondage calés campagne RP 2008
Minimum	-0.83333333	-0.58333333
25%	-0.0769231	-0.3577236
Médiane	0	-0.0487805
75%	0.0588235	0.1875000
Maximum	1.0869565217	7.2608696

Comparaison des solutions envisagées (3)

Conclusion : les allocations calculées à l'aide des poids calés sur la campagne RP N-5 peuvent être très éloignées des allocations théoriques

- Les calages des ZAE pour les campagnes RP 2008 et 2013 ont mené à des poids calés parfois très variables
- Paramètres de calage différents entre les campagnes RP 2008 et 2013 :
 - choix de bornes LO et UP différentes
 - variables de calage différentes
- Impact des franchissements de seuil de certaines communes

Méthode de calcul anticipé d'allocations

- Décision : calculer les allocations pour les enquêtes du 2nd semestre 2015, tirées dans la campagne RP 2014 :
 - en utilisant les effectifs par ZAE recensés dans la campagne N-5 = 2009
 - et les poids de sondage initiaux des ZAE
- Comme avant l'entrée en vigueur des NCEE : borner les allocations afin de ne pas sous/surcharger des enquêteurs

Tirage des échantillons

- Enquêtes du 1er semestre 2015 : tirage des échantillons
 - en mai/juin 2014
 - dans la campagne RP 2013
 - allocations directes du tirage, minimisant la dispersion des poids de sondage des logements sous contraintes (i.e. utilisant les poids des ZAE calés sur le RP 2013)

- Enquêtes du 2nd semestre 2015 : tirage des échantillons
 - en mai/juin 2015
 - dans la campagne RP 2014
 - allocations anticipées (calculées en mai/juin 2014 dans la campagne RP 2009, en utilisant les poids de sondage initiaux des ZAE) = paramètres exogènes du tirage

La nouvelle planification de l'échantillonnage

Merci de votre attention !

Contact

M. Pierre-Arnaud Pendoli

Tél. : 01 41 17 50 57

Courriel : pierre-arnaud.pendoli@insee.fr

Insee

18 bd Adolphe-Pinard
75675 Paris Cedex 14

www.insee.fr  

Informations statistiques :

www.insee.fr / Contacter l'Insee

09 72 72 4000

(coût d'un appel local)

du lundi au vendredi de 9h00 à 17h00