# Tirage coordonné d'échantillons : une application à l'Échantillon-Maître Nautile et à l'enquête Emploi

**Thomas Merly-Alpa**, Ludovic Vincent, Laurent Costa, Clément Guillo, Nicolas Paliod, Martin Chevalier, Thomas Deroyon

10ème Colloque Francophone sur les sondages

24 octobre 2018



- 1 La coordination avec l'enquête Emploi L'enquête Emploi Pourquoi coordonner les tirages?
- 2 Méthode de coordination Les Unités de Coordination Tirage direct ou indirect? Conserver les équilibrages
- 3 Calcul des probabilités de sélection Allocations régionales Probabilités de sélection des secteurs

#### Plan

- 1 La coordination avec l'enquête Emploi
- 2 Méthode de coordination
- 3 Calcul des probabilités de sélection

## Rappel : problématique et contexte

- Renouveler l'Échantillon-Maître (EM)
  - Une unité primaire est un regroupement de communes ;
  - C'est l'ensemble des logements dans lesquels se concentrera la collecte des enquêtes ménages;
  - Le nouvel EM sera mis en place en 2020 : tirage spatialement équilibré.
- Renouveler l'échantillon de l'enquête Emploi en continu (EEC)
  - Le nouvel échantillon de l'EEC débutera en juillet 2019.
- Coordination des deux échantillons?

## Quelques rappels sur l'Enquête Emploi en Continu

- L'enquête Emploi en Continu (EEC) est le pendant français de la *Labour Force Survey*.
- Elle permet de mesurer le taux de chômage, le taux d'activité ainsi que d'autres éléments sur l'emploi (temps partiel, ...).
- L'échantillon est « aréolaire » : il n'est pas issu directement d'un tirage de logements, mais d'un tirage de groupes de logements proches.

## Quelques rappels sur l'Enquête Emploi en Continu

- Ces groupes, ou « grappes », sont des ensembles géographiquement compacts d'une vingtaine de logements.
- L'échantillon est construit selon une logique ascendante : construction de grappes, puis de secteurs comportant 6 ou 7 grappes.
- La collecte se fait au sein d'un secteur pour assurer un effet de pseudo-panel sur la mesure de l'évolution des taux de chômage.

## Le tirage des secteurs

- Tirage spatialement équilibré des secteurs.
- Tirage stratifié car il existe des contraintes de précision régionale.
- Variables d'équilibrages socio-démographiques et liées aux revenus : proxy de l'activité et du chômage.

## Pourquoi coordonner les tirages?

- Car les échantillons des enquêtes ménages et de l'enquête Emploi sont renouvelés à moins d'un an d'intervalle.
- Car cela limite les déplacements d'enquêteur dans des secteurs Emploi isolés.

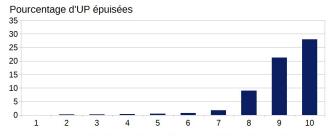
#### Plan

- 1 La coordination avec l'enquête Emploi
- 2 Méthode de coordination
- 3 Calcul des probabilités de sélection

## Tirer les secteurs dans les UP : épuisement des UP

- Tirer les secteur dans les UP?
  - Certaines UP s'épuisent trop vite à cause de la non-réinterrogation des ménages;
  - L'effet grappe est trop marqué : perte de précision importante.

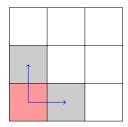
#### Épuisement des UP en cas de tirage de l'EEC dans les UP



Année après la première utilisation de l'échantillon-maître

## Quelle solution?

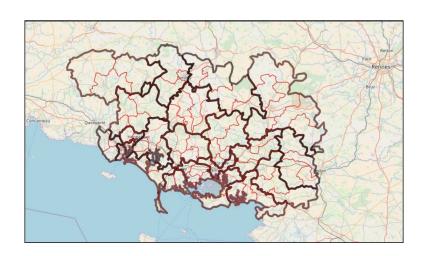
- Contrainte de la coordination : les secteurs emploi doivent être à proximité des unités primaires tirées.
- Tirer dans les unités primaires voisines :
  - Permet de garantir un tirage proche sans épuisement ni effet grappe trop fort;
  - Problèmes : comment calculer les probabilités de sélection ?
    Comment gérer les frontières ?



## Pourquoi constituer de nouvelles zones?

- Solution : dessiner de nouvelles zones (les unités de coordination - UC) en regroupant des unités primaires - UP
- Regrouper combien d'UP? Combien de logements?
- On décide d'un seuil minimal de 10 000 logements par UC :
  - Épuisement moindre des UP tirées;
  - Le tirage de l'EEC est plus précis avec plus de logements dans les UC;
  - Les UC restent peu étendues.

# Exemple du Morbihan



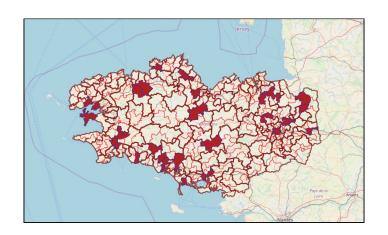
## Tirage direct des UC

- Méthode intuitive : tirage à 2 degrés
  - Tirage des UC en premier lieu
  - Tirage des UP et des secteurs dans les UC tirées
- Tirer 1 UP par UC tirée entraîne une moindre qualité de l'échantillon-maître.
- Ne pas contraindre le tirage d'1 UP par 1 UC limite la coordination.
- Cette solution n'est pas retenue!

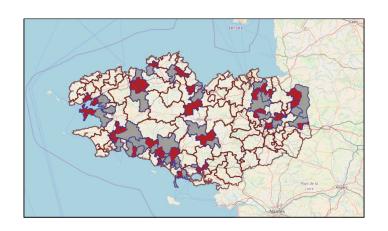
## Tirage indirect des UC

- Le problème du tirage direct des UC vient du trop grand nombre de contraintes sur le tirage des UP.
- Le tirage des secteurs pour l'EEC aurait fonctionné dans la méthode de tirage direct.
- Idée : Tirer les UP en premier pour ne pas avoir la contrainte d'une unité de coordination par UC.
- Tirage indirect des UC :
  - Étape 1 : Tirage des UP;
  - Étape 2 : Récupération des UC associées aux UP tirées;
  - Étape 3 : Tirage des secteurs dans les UC tirées.

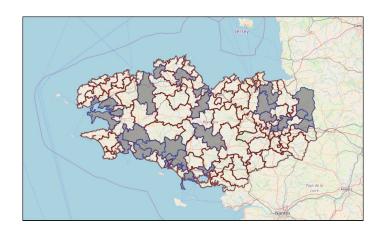
Étape 1 : Tirage des UP



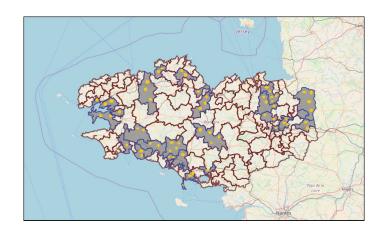
# Étape 2 : Récupération des UC



# Étape 2 : Récupération des UC



# Étape 3 : Tirage des secteurs



## L'équilibrage indirect des UP

- Question restante : comment tirer les UP pour qu'elles soient précises pour l'EM et pour que les UC soient précises sur l'EEC?
- Les UP sont équilibrées sur des variables d'intérêt des enquêtes ménages.
- Les UP déterminent indirectement les secteurs qui pourront être tirés.
- Nécessité d'intégrer des variables dans le tirage des UP pour que les UC soient tirées de manière précise sur les variables de l'EEC.

#### Les variables transformées

- Création de variables transformées pour intégrer dans l'équilibrage des UP des variables de l'EEC qui permettront aux UC tirées d'être précises sur ces variables.
- On se place dans le contexte de la méthode généralisée de partage des poids.
- Les UC j(i) sont tirées indirectement par leurs UP i.
- On pondère le lien  $\theta_{i,j(i)}$  par le nombre de logements, proportionnel à la probabilité de sélection.

#### Les variables transformées

• Pour chaque variable d'équilibrage X au niveau UC, on construit les variables associées  $\tilde{X}_i$  au niveau UP :

$$\tilde{X}_i = \theta_{i,j(i)} X_{j(i)}$$

• Alors, si l'échantillon d'UP est équilibré sur  $\tilde{X}$ , on a un échantillon d'UC équilibré sur X :

$$\sum_{j \in S_{UC}} w_j^{UC} X_j = \sum_{j \in S_{UC}} \sum_{i | j(i) = j} \theta_{i,j} w_i^{UP} X_j = \sum_{i \in S_{UP}} w_i^{UP} \tilde{X}_i$$

#### Plan

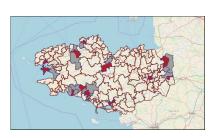
- 1 La coordination avec l'enquête Emploi
- 2 Méthode de coordination
- 3 Calcul des probabilités de sélection

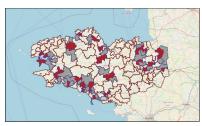
#### Choix des allocations

Les allocations régionales concernent deux éléments :

- le nombre d'UP tirées par région ;
- le nombre de secteurs Emploi tirés par région.

## Tirer plus d'UP ou de secteurs?





Tirage de 15 UP

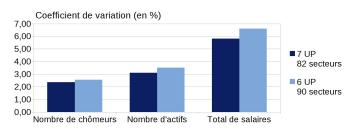
Tirage de 29 UP

	+ d'UP	+ de secteurs
précision EM	+	neutre
précision EEC	+	+

Pour la précision de l'EEC, quel arbitrage entre UP et secteurs?

## Tirer plus d'UP ou de secteurs?

# Influence des paramètres sur la précision de l'enquête Emploi dans le Limousin



- Faible effet d'une augmentation de 10% de secteurs;
- Fort effet d'une diminution de 10% d'UP.

## Calcul de la probabilité de tirage des secteurs

- Comment calculer  $\pi_k$  pour un secteur k?
- Objectifs : respecter les allocations tout en assurant l'équi-pondération.
- On calcule la probabilité de sélection conditionnelle à la sélection de l'UC i(k) de la façon suivante :

$$\pi_k = \frac{n - n_{\text{exh}}}{(NbUC - NbUC_{\text{exh}})NbSect_{i(k)}}$$

où n est l'allocation de l'EEC, NbUC le nombre d'UC tirées et  $NbSect_{i(k)}$  le nombre total de secteurs dans l'UC i(k).

• Les UC exhaustives sont traitées à part, directement.

#### Pour conclure

#### Pour la première fois, EM et échantillon de l'EEC sont coordonnés :

- Facilite la collecte
  - Optimisation de l'étendue des grappes et secteurs de l'EEC, et des unités primaires de l'EM
  - Constitution des unités de coordination (UC) regroupant les UP tirées et les secteurs
- En conservant une bonne précision
  - Abandon du recensement de la population pour l'EM
  - Amélioration des variables d'équilibrage mobilisées
  - Apport de l'équilibrage spatial
  - Équilibrage indirect des UC via l'équilibrage des UP

Merci pour votre attention!

## Choix de la méthode de tirage

- Utilisation d'un tirage spatialement équilibré à probabilités inégales
- L'équilibrage permet d'être plus précis sur les variables d'intérêts corrélées aux variables d'équilibrage
- le tirage spatial permet de :
  - limiter la corrélation spatiale
  - couvrir l'espace national

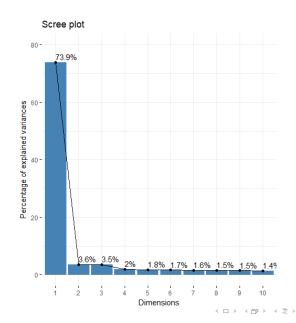
## Les variables d'équilibrage

- Grand nombre de variables socio démographiques disponibles, peut-on toutes les utiliser dans l'équilibrage?
- Trop de contraintes, ce n'est pas envisageable, perte de précision sur les variables ne participant pas à l'équilibrage
- On essaye de choisir quelques variables représentatives de l'ensemble des variables (à dire d'expert)
- Solution non retenue, on ne prend pas en compte assez d'information auxiliaire

## Les variables d'équilibrage

- Utilisation des méthodes d'analyse factorielle pour réduire le nombre de dimensions du problème
- Étude des corrélations entre les variables pour éviter la redondance dans l'information utilisée

# Utilisation de l'ACP : inertie expliquée par les premiers axes



#### Méthode hybride

- Les axes d'ACP sont par construction très explicatifs
- L'ACP permet de réduire la dimension du problème tout en gardant un pouvoir explicatif suffisamment élevé
- Méthode retenue, on veut néanmoins s'assurer d'un équilibrage parfait pour certaines variables (exemple : probabilités d'inclusion, revenu)
- On équilibre donc à la fois sur les axes d'ACP et sur ces variables cibles

#### Paramètres

- Participation des variables cibles à l'ACP
- Ordre des variables
- ⇒ Pour le tirage des UP, on retiendra donc la méthode hybride combinant ACP et équilibrage direct, en plaçant les axes en première position