

Constitution de l'échantillon maître pour les extensions régionales : une procédure de tirage d'échantillons équilibrés emboîtés

Marc Christine, Emmanuel Gros
Insee
JMS 2009





1. Contexte

- Système échantillonnage des enquêtes ménages : sondage à deux degrés reposant sur l'échantillon-maître.
- L'EM n'est représentatif qu'au niveau national.
 - ➔ Constitution d'une base de sondage complémentaire de l'EM : l'échantillon-maître pour les extensions régionales (EMEX).
- En cas de demande d'extensions régionales, tirage des logements au sein de l'ensemble EM+EMEX.



1. Problématique : tirage de l'EM et de l'EMEX

- Objectif : définir un nouveau système de tirage de l'EM et de l'EMEX, assurant les trois propriétés suivantes :
 - ✓ ZAE tirées avec des probabilités proportionnelles à la taille...
 - ✓ ... en respectant les mêmes probabilités pour les ZAE-EM que celles que l'on aurait imposées si l'on avait tiré seulement l'EM...
 - ✓ ...avec des conditions d'équilibrage régional **assurées pour chaque ensemble** : EM+EMEX et EM seul.



1. Retour sur le tirage de l'EMEX de 2002

- En 2002, le tirage de l'EMEX avait été postérieur à celui de l'EM → EMEX constitué conditionnellement au tirage de l'EM.
- Impossible alors de garantir toutes les propriétés voulues → La condition d'équilibrage avait été abandonnée au profit du respect des probabilités d'inclusions proportionnelles à la taille.
- Aujourd'hui, constitution simultanée de l'EM et de l'EMEX → permet de respecter les différentes contraintes, et même d'envisager un raffinement du concept d'EMEX (*EMEX restreint*, *EMEX élargi*)



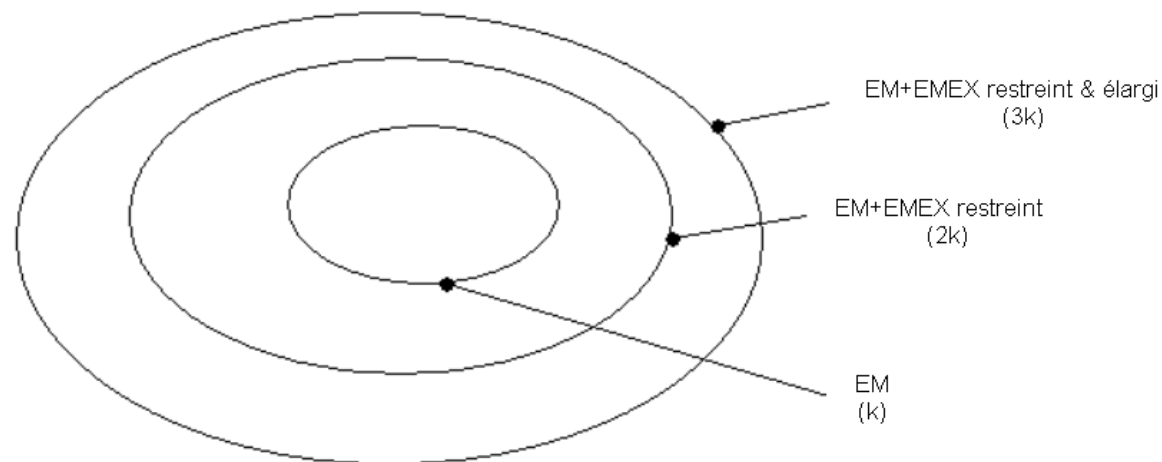
1. Le tirage simultané de l'EM et de l'EMEX

- Un tirage simultané EM – EMEX dans chaque région...
- ... qui respecte à la fois les contraintes d'équilibrage et de probabilités d'inclusion.
- Extension du concept d'EMEX :
 - ✓ un EMEX restreint mobilisé pour des extensions de taille modérée : doublement de l'effectif des ZAE-EM ;
 - ✓ si nécessaire, pour les grandes extensions : mobilisation supplémentaire d'un EMEX élargi, qui permet de tripler l'effectif des ZAE-EM.



2. Méthodologie – principes

- Tirage en plusieurs phases emboîtées : EM+EMEX restreint et élargi, EM+EMEX standard, puis EM.
- On commence par mettre de côté les unités exhaustives. L'univers $U = \text{complément}$.



k = nombre d'unités non exhaustives à tirer pour l'EM.



2. Méthodologie – tirage en deux phases

- 1ère phase : échantillon s_1 (EM+EMEX) de taille $2k$, probabilités d'inclusion π_i^1
 - 2ème phase : conditionnellement au tirage de s_1 , échantillon s_2 (EM seul) de taille k , probabilités d'inclusion $\pi_i^{2/1}$
- ➔ Quatre paramètres à déterminer – probabilités d'inclusion et variables d'équilibrage de 1^{ère} et 2^{nde} phase – à l'aide des quatre contraintes suivantes :
- ✓ conditions d'équilibrage ;
 - ✓ contrainte de taille fixe des échantillons ;
 - ✓ respect des probabilités d'inclusions PPT ;
 - ✓ probabilités d'inclusion $\in [0,1]$.

2. Conditions d'équilibrage sur la variable X

➤ Pour l'ensemble EM+EMEX :
$$\sum_{i \in s_1} \frac{X_i}{\pi_i^1} = \sum_{i \in U} X_i$$

➤ Pour l'EM seul :
$$\sum_{i \in s_2} \frac{X_i}{\pi_i^1 \pi_i^{2/1}} = \sum_{i \in U} X_i$$

D'où :
$$\sum_{i \in s_2} \frac{X_i / \pi_i^1}{\pi_i^{2/1}} = \sum_{i \in s_1} (X_i / \pi_i^1)$$

➔ Au sein du 1^{er} échantillon, le 2nd échantillon (tiré conditionnellement au 1^{er}) doit être **équilibré sur la variable** X / π^1 .



2. Contraintes de taille fixe

➤ Pour l'ensemble EM+EMEX :

$$\sum_{i \in U} \pi_i^1 = 2k$$

➤ Pour l'EM seul :

$$\sum_{i \in S_1} \pi_i^{2/1} = k$$



2. Respect des probabilités d'inclusion

- *Respect des probabilités d'inclusion* des ZAE de l'EM seul :

$$\pi_i^1 \pi_i^{2/1} = k \frac{N_i}{N - N^{exh}}$$

D'où :

$$\pi_i^{2/1} = k \frac{1}{\pi_i^1} \frac{N_i}{N - N^{exh}} \quad (*)$$

Soit :

$$\boxed{\sum_{i \in s_1} \frac{N_i}{\pi_i^1} = N - N^{exh}}$$

➔ **Le 1^{er} échantillon doit être équilibré sur la taille des unités.**



2. Cohérence des probabilités d'inclusion

- Reste à *déterminer les probabilités d'inclusion du 1^{er} échantillon*, qui doivent vérifier :

$$\sum_{i \in U} \pi_i^1 = 2k, \pi_i^1 \in [0,1] \text{ et } \pi_i^{2/1} \in [0,1]$$

La dernière contrainte injectée dans (*) implique

$$k \frac{N_i}{N - N^{exh}} \leq \pi_i^1 \leq 1, \text{ équation admissible.}$$



2. Cohérence des probabilités d'inclusion

- Une solution naturelle pour déterminer les π_i^1 :

$$\pi_i^1 = 2k \frac{N_i}{N - N^{exh}} \quad . \quad \text{Problème : peut être } > 1$$

- ➔ Seuil d'exhaustivité spécifique à l'ensemble EM+EMEX :

$$S_1 = \frac{N - N^{exh}}{2k}$$

- D'où k_1 nouvelles unités exhaustives pour l'ensemble EM+EMEX et :

$$\pi_i^1 = (2k - k_1) \frac{N_i}{N - N^{exh} - N_1^{exh}}$$

3. Impact sur la qualité des estimations réalisées à partir de l'EM



- En théorie :
 - ✓ les probabilités d'inclusion dans l'EM étant inchangées, l'impact en terme de biais devrait être nul ;
 - ✓ le fait de procéder à un sondage en deux phases rajoute un niveau d'aléa supplémentaire au processus de tirage de l'EM, ce qui peut potentiellement dégrader la précision des estimations.

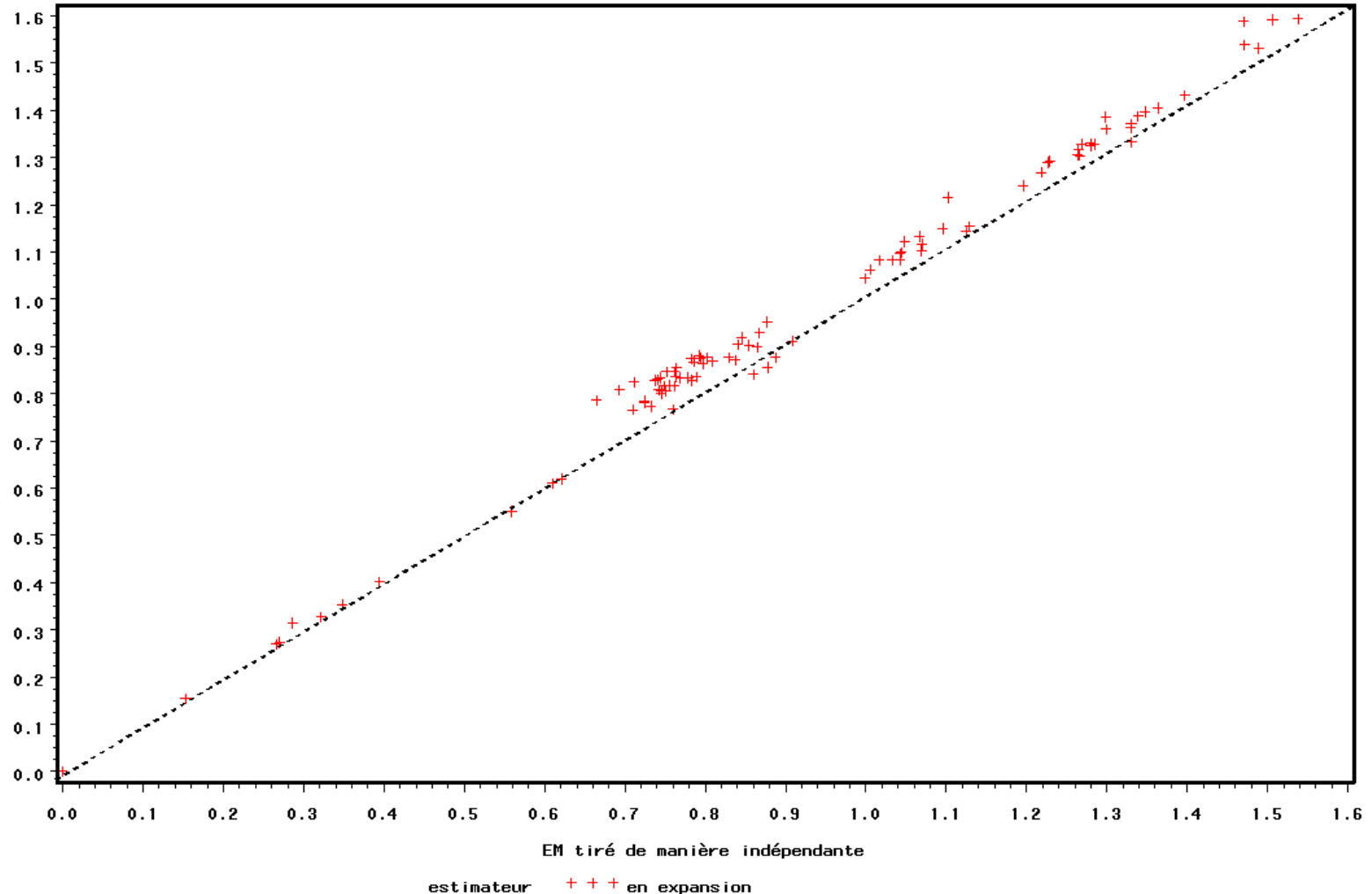
- En pratique : simulations à partir des données RP99.
 - ➔ aucune influence sur le biais des estimations
 - ➔ impact mesuré en terme de précision



3. Un impact très limité en terme de précision

Comparaison de la précision des estimations (CV) suivant le mode de tirage de l'EM

EM tiré en deux phases





4. En résumé (1)

- Pour assurer un équilibrage sur le total d'une variable X , pour l'EM seul et pour l'ensemble EM+EMEX, il suffit de tirer :
 - ✓ un premier échantillon s_1 , équilibré sur X , avec des probabilités d'inclusion π_i^1 , qui représentera l'ensemble EM+EMEX ;
 - ✓ puis, au sein de s_1 et conditionnellement au tirage de ce dernier, un échantillon s_2 , équilibré sur la variable X / π^1 , avec des probabilités d'inclusion $\pi_i^{2/1}$, qui représentera l'EM seul.



4. En résumé (2)

- Des ZAE exhaustives spécifiques EMEX (retenues d'office, si l'on mobilise l'EMEX, mais seulement avec une certaine probabilité si l'on n'utilise que l'EM).
- La procédure se généralise pour le tirage de plus de deux échantillons → procédure en trois phases pour l'EM, l'EMEX standard et l'EMEX élargi.
- Un impact très limité sur la qualité des estimations :
 - ✓ aucun impact sur le biais des estimations ;
 - ✓ une légère perte de précision.