

Méthodologie d'échantillonnage de la future Enquête Emploi en Continu à Mayotte

JOURNÉES DE MÉTHODOLOGIE STATISTIQUE DE L'INSEE

Julien Jamme (Insee, DMCSI)
Cyril Favre-Martinoz (Insee, DSDD)
Édouard Fabre (Insee, Criem)
Joachim Timoteo (Insee, Criem)

31/03/2022



- 1 CONTEXTE ET ENJEUX
- 2 ÉCHANTILLONNAGE
- 3 RÉSULTATS DES SIMULATIONS
- 4 CONCLUSION

Depuis 2013, l'Insee organise une enquête Emploi annuelle à Mayotte.

- Vieillissement de l'enquête actuelle ;
- Une dérogation européenne qui arrive à échéance.

⇒ **passage à une enquête trimestrielle en continu**

Point de départ = Méthodologie appliquée dans les autres DOM en 2013 :

- 1 Partition du territoire en Secteurs d'Activité des Enquêteurs ;
- 2 Tirage systématique de logements à partir du RP.

OBJECTIFS

- Proposer une méthode d'échantillonnage de l'EEC à Mayotte ;
- Estimer la précision d'estimateurs transversaux.

CARACTÉRISTIQUES DE L'ENQUÊTE EMPLOI

- Une enquête trimestrielle **en continu** par **panel rotatif**.
- Population Cible : Individus de 15-89 ans en logement ordinaire.
- Unités échantillonnées : Logements
- Unités statistiques : Individus

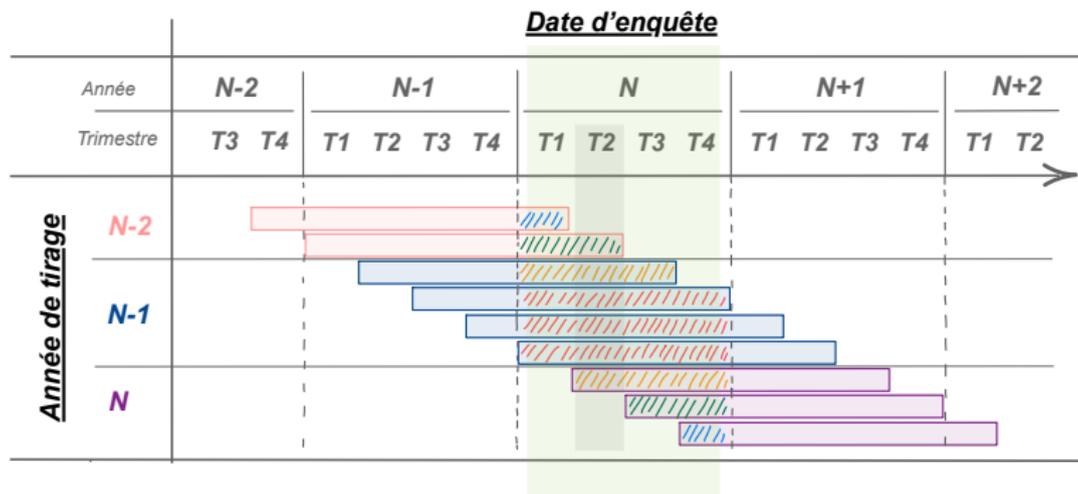


FIGURE – Roulement des échantillons de l'EEC

LA BASE DE SONDAGE

UNE BASE D'ADRESSES

- Une partie importante du parc de logements est précaire ;
- Besoin d'une information suffisamment fiable et fraîche
- Absence de base de sondage de logements fiable et suffisamment récente à Mayotte ;
- Recours à une base d'adresses, la base cartographique :
 - ▶ issue d'une enquête cartographique ;
 - ▶ un cinquième des îlots est enquêté chaque année ;
 - ▶ répertorie l'ensemble des adresses situés dans ces îlots ;
 - ▶ pauvre en informations auxiliaires.

LA CONSTRUCTION DE SECTEURS D'ACTIVITÉ DES ENQUÊTEURS (SAE)

Pourquoi ?

- Pour répartir l'échantillon sur l'ensemble du territoire ;
- Pour répartir le plus également possible la charge de l'enquête entre enquêteurs : un SAE pris en charge par un enquêteur ;
- Pour assurer la continuité de l'enquête : 26 SAE, soit 2 SAE enquêtés chaque semaine d'un trimestre (en année non bissextile).

Comment ?

- Regroupements d'îlots par l'algorithme du voyageur de commerce ;
- Critère d'arrêt : seuil fixé de logements par groupe de rotation ;
- Nombreuses itérations nécessaires ;
- Le choix se porte sur le jeu de secteurs minimisant :
 - ① l'écart au seuil dans chaque SAE et pour chaque groupe de rotation ;
 - ② l'étendue spatiale des secteurs ;
 - ③ les temps de collecte estimés.
- Contrainte de contiguïté relâchée.

CONTRAINTES ET OBJECTIFS D'ÉCHANTILLONNAGE

① Contraintes institutionnelles ;

- ▶ Précision européenne requise ;
- ▶ Choix Insee : respecter la contrainte européenne à coût minimum.

⇒ Tirer un échantillon de **936 logements entrants par an** ;

② Contraintes d'organisation de la collecte

- ▶ Répartir la collecte sur l'ensemble des semaines de l'année ;

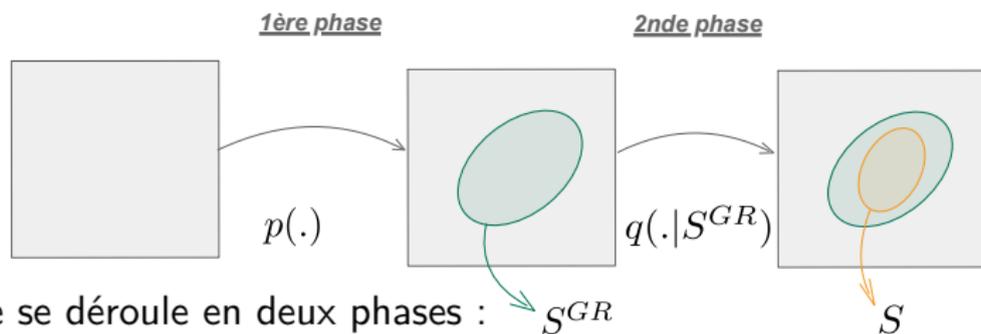
⇒ Obtenir **une égale répartition de l'échantillon** entre les SAE, soit **36 ± 4 logements entrants chaque année**

③ Contraintes méthodologiques

- ▶ Absence de base de sondage de logements de bonne qualité ;
- ▶ **Autopondération** du plan de sondage ;
- ▶ Obtenir la **meilleure précision** possible pour les estimateurs étudiés.

ÉCHANTILLONNAGE

UN TIRAGE EN DEUX PHASES



Le tirage se déroule en deux phases :

- Phase 1** : Tirage des groupes de rotation d'îlots
 - ▶ Méthodologie fixée et tirage effectué en 2020 ;
 - ▶ L'échantillonnage des logements ne se fera que dans un cinquième des îlots chaque année.
- Phase 2** : Tirage des logements conditionnellement au groupe de rotation d'îlots tiré en phase 1.
 - ▶ Tirage à deux degrés (adresses puis logements) ;
 - ▶ Phase sur laquelle porte notre travail.
- La présence de **non-réponse totale** ajoute en réalité une troisième phase de tirage.

ÉCHANTILLONNAGE

LE POINT DE DÉPART

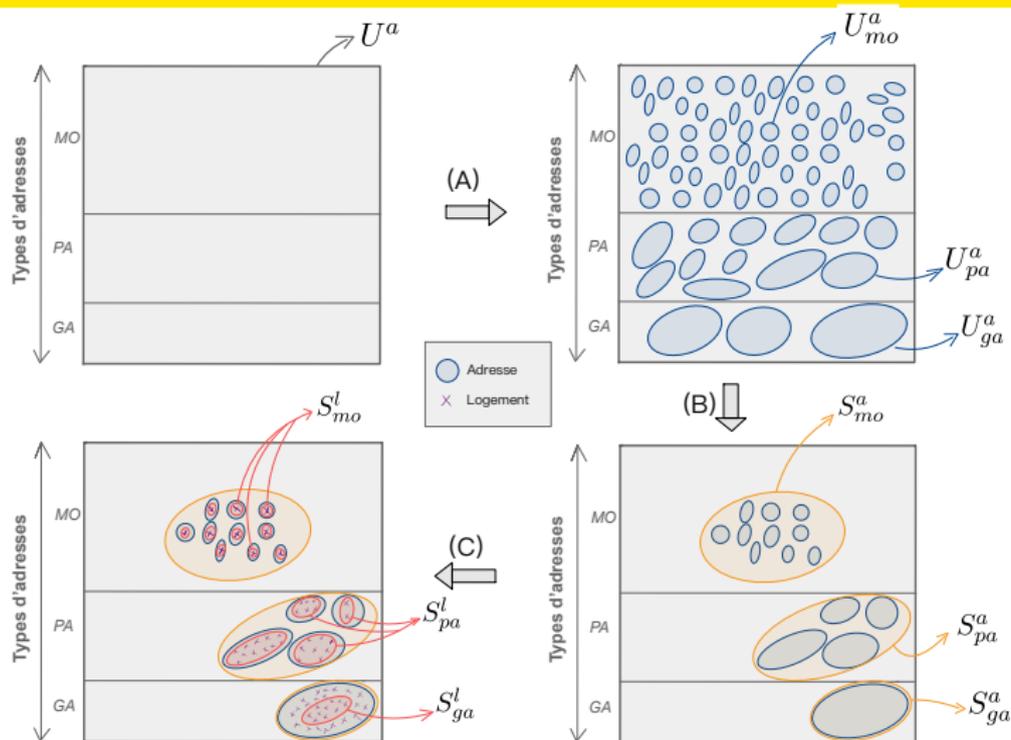


FIGURE – Un scénario de base pour servir de point de référence

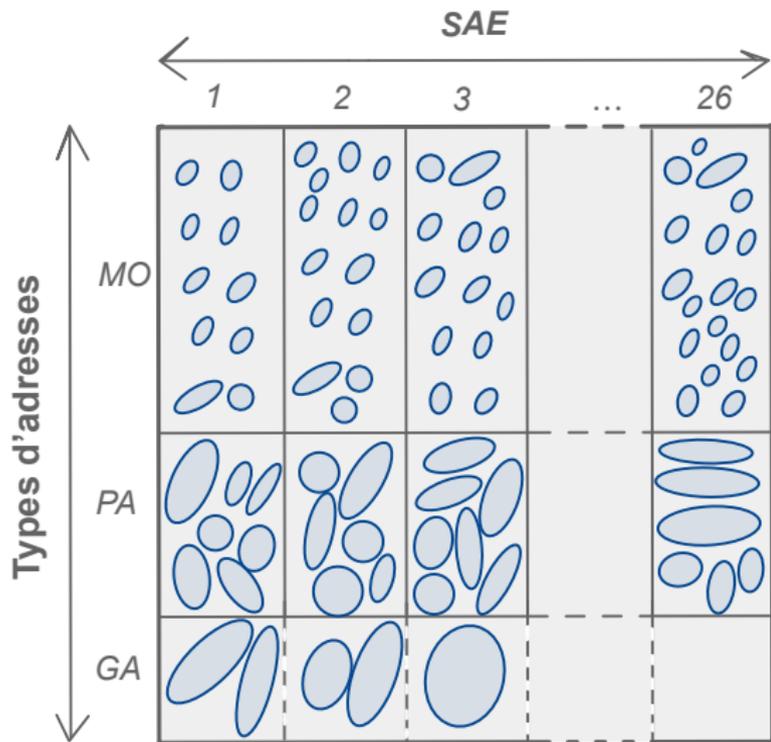


FIGURE – Les adresses par type et SAE

ÉCHANTILLONNAGE

LE SCÉNARIO PROPOSÉ

DIFFICULTÉ PRINCIPALE RENCONTRÉE

Toute tentative de mieux répartir l'échantillon nuit en particulier à l'autopondération du plan de sondage.

IDÉE CENTRALE DU PLAN PROPOSÉ

Assurer la répartition par secteur (SAE) **en fixant d'emblée une allocation égale par SAE** et effectuer **un tirage stratifié par type d'adresses *indépendamment* dans chacun des SAE.**

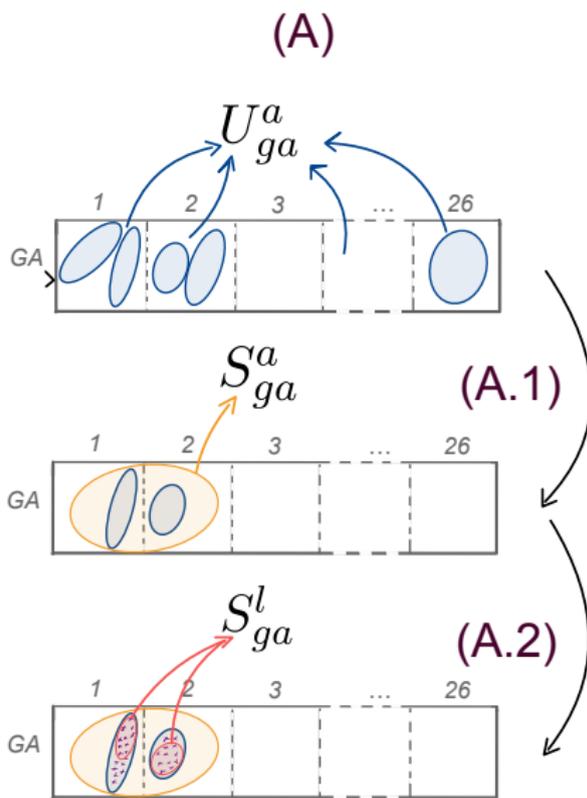


FIGURE – Étape A : Tirage préalable des logements en grandes adresses (Tirage à 2 degrés). L'allocation de logements est proportionnelle à la taille de la strate.

$$\begin{array}{ccc} \text{(A.3)} & & \Downarrow \\ n_{1,ga}^l = 10, n_{2,ga}^l = 10, n_{3,ga}^l = 0 \dots n_{26,ga}^l = 0, & & \\ \text{(A.4)} & & \Downarrow \\ {}^{(1)}\tilde{n}_1^l = 26, {}^{(1)}\tilde{n}_2^l = 26, {}^{(0)}\tilde{n}_3^l = 36 \dots {}^{(0)}\tilde{n}_{26}^l = 36 & & \end{array}$$

FIGURE – Ajustement des allocations par SAE en fonction du nombre de logements en GA tirés dans chacun des SAE (ici avec un seuil de 10 logements entre PA et GA)

ÉCHANTILLONNAGE

LE SCÉNARIO PROPOSÉ

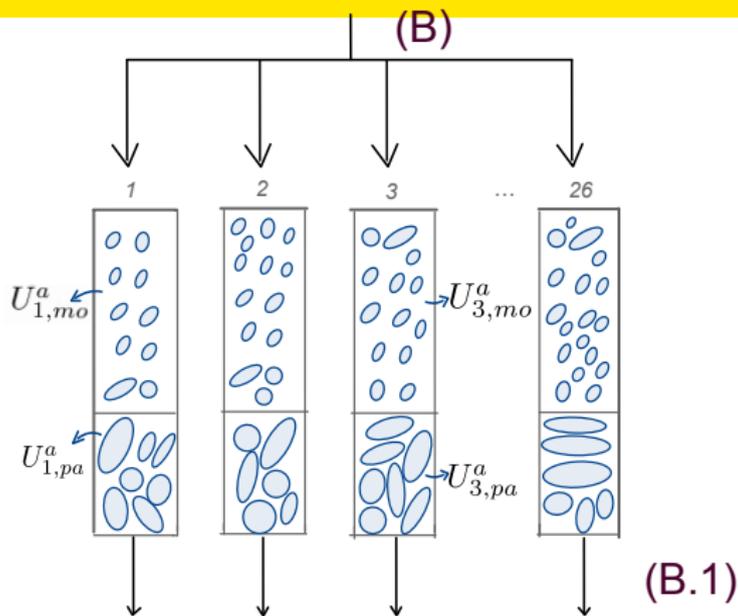


FIGURE – B.1 : Pour le tirage des monologements (MONO) et des logements en petites adresses (PA), chaque secteur (SAE) est traité indépendamment. Au départ, à chacun est alloué le même nombre de logements (36).

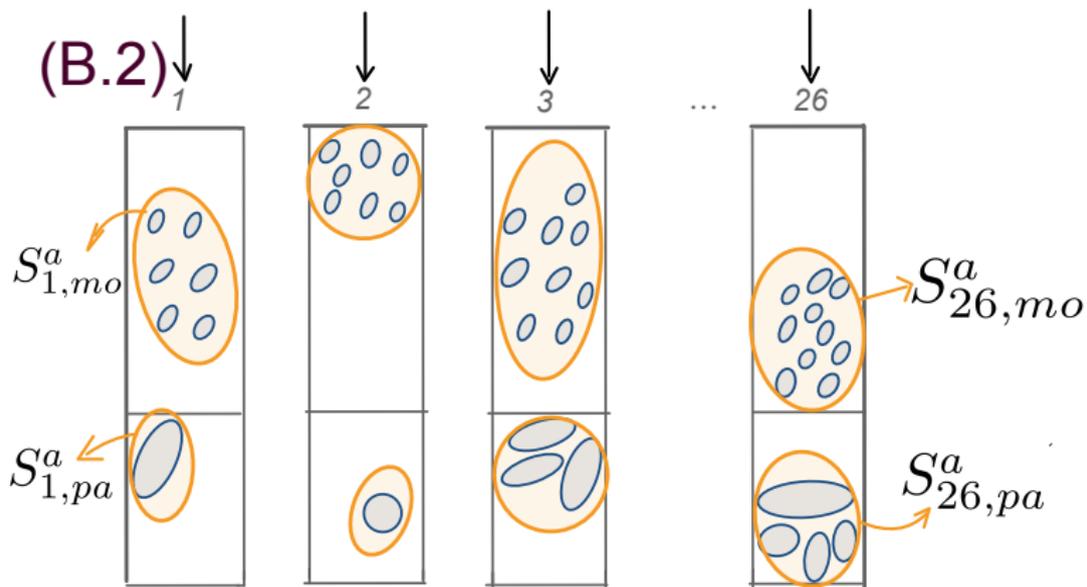


FIGURE – B.2 : On récupère les allocations ajustées suite au tirage des logements en GA puis on tire les logements par un tirage stratifié par type d'adresses.

- Données mobilisées : recensement de la population 2017 (exhaustif) ;
- Simulations par la méthode de Monte-Carlo.

Le processus complet d'échantillonnage reproduit :

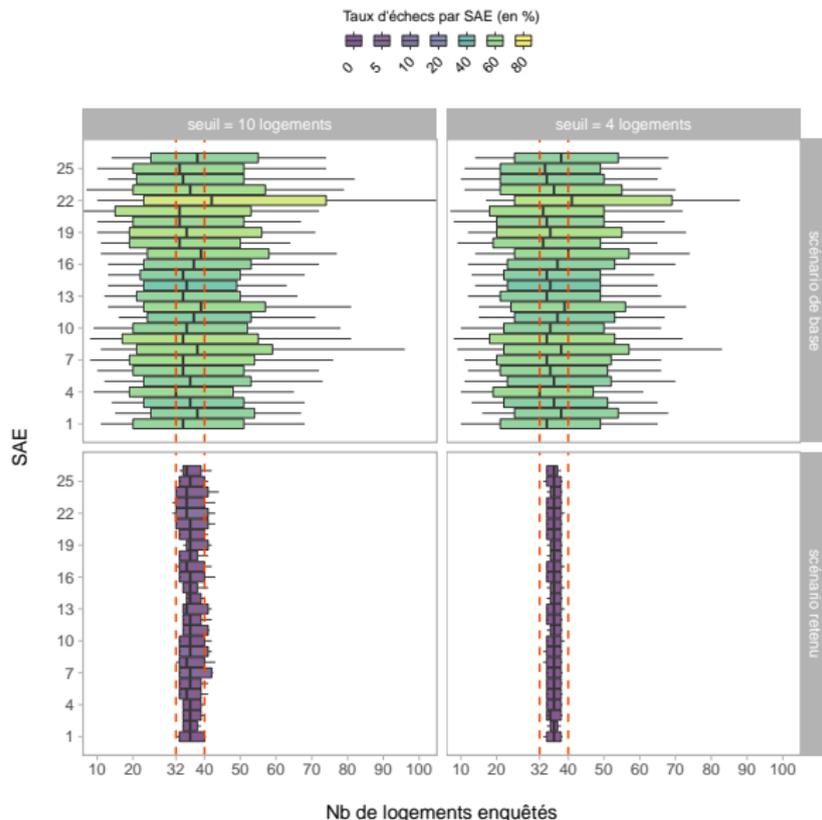
- 1 Phase 1 : Tirage des groupes de rotation des îlots ;
- 2 Construction de SAE optimisés sur les groupes de rotation constitués ;
- 3 Phase 2 : Tirage des logements selon le scénario retenu (ou à l'occasion le scénario de base) ;
- 4 Phase 3 : Tirage des répondants - en appliquant un mécanisme de réponse aléatoire basé sur des Groupes Homogènes de Réponse ;

⇒ Difficulté rencontrée :

- La reproduction de l'étape 2 impossible dans les limites temporelles de ce travail ;
- En cause : des temps de calculs trop longs.

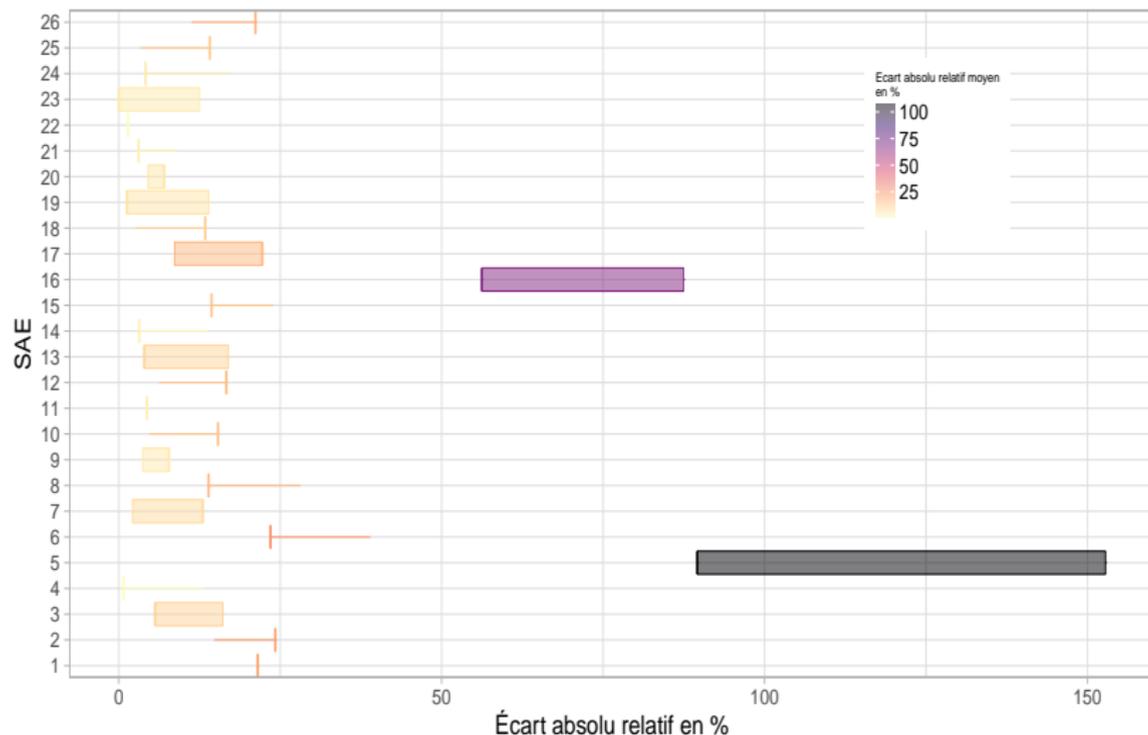
RESPECT DES CONTRAINTES

RÉPARTITION DE L'ÉCHANTILLON PAR SAE



RESPECT DES CONTRAINTES

VÉRIFICATION EMPIRIQUE DE L'AUTOPONDÉRATION



PRÉCISION DE L'ESTIMATEUR DU TAUX DE CHÔMAGE

QUEL ESTIMATEUR ?

DÉFINITION

Le taux de chômage C est le nombre de chômeurs BIT (t_y) rapporté au nombre d'actifs BIT (t_z) dans la population des 15 ans ou plus.

On estime C avec un estimateur par le ratio :

- Chaque membre du ratio est estimé par :
 - ▶ un estimateur par expansion pour tenir compte de la phase 1 ;
 - ▶ un estimateur corrigé de la non-réponse totale (seconde expansion) ;
 - ▶ un estimateur calé sur les marges de variables auxiliaires (la population par sexe, âge, lieu de naissance, etc.) disponibles avec le RP 2017.
- Le calage est d'autant plus efficace ici que les marges sont calculées à partir de la même source qui sert à tirer l'échantillon dans nos simulations (RP 2017).

PRÉCISION DE L'ESTIMATEUR DU TAUX DE CHÔMAGE

ESTIMATION DIRECTE VS ESTIMATION INDIRECTE

Pour faire face à la difficulté de reproduire le processus complet de l'échantillonnage, on s'appuie sur un scénario de base indifférent à la construction des SAE (tirage à deux degrés stratifié par type d'adresses).

1. Estimation directe à partir du scénario retenu :

- ▶ Reproduction du processus complet sans optimisation des SAE ;
- ▶ Autopondération du plan retenu moins bien respectée ;
- ▶ Surestimation de la variance ;

2. Estimation partielle de phase 2 :

- ▶ On raisonne conditionnellement à la première phase.

PRÉCISION DE L'ESTIMATEUR DU TAUX DE CHÔMAGE

ESTIMATION DIRECTE VS ESTIMATION INDIRECTE

TABLE 5.2 – Précision globale de l'estimation du taux de chômage

	Estimation	Biais rel.	Écart-type	CV	Précision à 95%	
	(en %)	(en %)	(en pts de %)	(en %)	trim.	annuelle
Scénario de base	27,70	0,13	1,91	6,91	3,82	3,41
Scénario retenu	27,73	0,24	1,97	7,11	3,93	3,47

TABLE 5.3 – Précision de phase 2 de l'estimation du taux de chômage

	Estimation	Biais rel.	Écart-type	CV	Précision à 95%	Effet de plan
	(en %)	(en %)	(en pts de %)	(en %)	trim.	
Scénario de base	27,75	0,31	1,85	6,70	3,71	
Scénario retenu	27,74	0,29	1,78	6,42	3,55	

3. Estimation indirecte à partir du scénario retenu et du scénario de base :
- ▶ On fait l'hypothèse que le rapport des variances entre processus complet et processus conditionné à la phase 1 est identique pour les deux scénarios, c'est-à-dire :

$$\frac{\mathbb{V}_{tot}^{ret}(\hat{C})}{\mathbb{V}_{P2}^{ret}(\hat{C})} = \frac{\mathbb{V}_{tot}^{base}(\hat{C})}{\mathbb{V}_{P2}^{base}(\hat{C})} \quad (1)$$

TABLE – Estimations directe et indirecte du taux de chômage avec le plan de sondage retenu

	Précision (en pts de %)	
	Est. directe	Est. indirecte
est. trimestrielle	3,93	3,66
est. annuelle	3,47	3,31

- Un taux de chômage annuel estimé à $\pm 3,3$ **points de %** ;
- Une estimation indirecte qui rend un plus justice au plan proposé ;

- Une précision dégradée par rapport aux ± 2 points de l'enquête actuelle ;
- Une enquête plus lourde et réalisée à coûts constants quasiment ;
- Avec un échantillon plus petit (1 400 vs 3 000) ;
- Un échantillon de 7 000 logements serait nécessaire pour obtenir la même précision annuelle ;
- Une méthode qui répond à l'ensemble des contraintes.

- Livraison d'un programme de tirage de l'échantillon prêt à l'emploi ;
- Echéance : T1-2024, soit un premier tirage fin 2021 début 2022 ;
- Des pistes d'amélioration :
 - ▶ Retoucher les deux SAE problématiques (pérennité de l'enquête) ;
 - ▶ Améliorer la typologie des îlots (légers gains de précision éventuels).
 - ▶ Pour le futur :
 - ★ Espérer la disponibilité d'une base de sondage de logements...
 - ★ plus riche en informations auxiliaires.

	Point de comparaison	Effet attendu sur la précision	Remarque
Stratification par type d'adresses	SRS	+	<i>Les GA sont atypiques par rapport aux variables d'intérêt</i>
Homogénéité de la taille des SAE	Scénario de base	-	<i>Logique d'organisation</i>
Dispersion spatiale de l'échantillon (SAE)	Scénario de base	++	<i>Un peu comme un équilibrage spatial Atténue l'effet de l'autocorrélation spatiale des phénomènes socio-économiques.</i>
Seuil à 4 logements	Seuil à 10 logements	++	<i>Effet de grappes (autocorrélation spatiale) est mieux contenu.</i>
Tirage systématique au 1 ^{er} degré sur variables triées	Scénario de base	+	<i>Effet d'autant plus fort que les variables de tri sont liées aux variables d'intérêt.</i>