

# TIRAGES (SPATIALEMENT) ÉQUILIBRÉS : PROPRIÉTÉS ET UTILISATIONS PRATIQUES

Julien Nicolas<sup>1</sup> & Nicolas Paliot<sup>2</sup>

<sup>1</sup> Insee, [julien.nicolas@insee.fr](mailto:julien.nicolas@insee.fr)

<sup>2</sup> Drees, [nicolas.paliot@sante.gouv.fr](mailto:nicolas.paliot@sante.gouv.fr)

**Résumé.** L'objectif de cette communication est de présenter différents types d'équilibrage pour le tirage d'unités primaires dans le contexte de tirage à deux degrés. Sur la base de l'exemple du tirage d'unités primaires pour l'enquête sociale européenne (ESS), sont présentés les niveaux de précisions obtenues avec le recours à des tirages équilibrés ou spatialement équilibrés. Enfin, un focus est réalisé pour voir comment il est possible de concilier des exigences de précision nationale et locales, notamment avec l'utilisation de tirages spatialement équilibrés stratifiés.

**Mots-clés.** Échantillonnage, tirage à deux degrés, unités primaires, échantillonnage équilibré, échantillonnage spatialement équilibré, échantillonnage stratifié

**Abstract.** This communication shows different types of balanced sampling of primary units in the context of two-stage sampling. Based on the example of the sampling of primary units for the European Social Survey (ESS), the levels of precision obtained using balanced or spatially balanced sampling are presented. Finally, a focus shows how it is possible to reconcile national and local precision, in particular through the use of stratified spatially balanced sampling.

**Keywords.** Sampling, two-degree sampling, primary units, balanced sampling, spatially balanced sampling, stratified balanced sampling

## 1 Introduction et cadre théorique

Dans le cadre des enquêtes ménages, les tirages équilibrés sont utilisés régulièrement à l'Insee, pour les tirages d'unités primaires, par exemple lors d'enquêtes ayant recours à une collecte en face-à-face. On réalise alors un tirage à deux degrés dont le premier degré consiste en un échantillon de zones de collecte, les unités primaires, avant d'effectuer un second tirage d'individus ou de logements au sein de ces unités primaires permettant ainsi de concentrer la collecte dans certaines zones.

### 1.1 Rappel théorique des tirages équilibrés et spatialement équilibrés

L'objectif des tirages équilibrés est de réduire la variance d'estimation sur des grandeurs associées à une ou plusieurs variables  $Y$  d'intérêt collectées au cours de l'enquête, en mobilisant des variables auxiliaires de la base de sondage.

Par exemple, l'objectif peut être de minimiser la variance  $V(\hat{t}_Y)$  de l'estimateur Horvitz-Thompson du total de  $Y$  défini ainsi :  $\hat{t}_Y = \sum_{k \in S} \frac{y_k}{\pi_k}$ . (1).

En supposant que  $Y$  est corrélée aux variables auxiliaires  $X_1, \dots, X_p$  de la base de sondage, on va mobiliser ces variables lors du tirage pour diminuer la variance  $V(\hat{t}_Y)$ .

Comme mentionné par Costa et Merly-Alpa (2017), un échantillon sera dit équilibré sur une ou plusieurs variables auxiliaires de la base de sondage, lorsque pour chacune d'entre elles, l'estimateur Horvitz-Thompson du total coïncide exactement<sup>1</sup> avec le vrai total issu de la base de sondage :

$$\forall j \in \{1 \dots p\}, \hat{t}_{X_j} = \sum_{k \in S} \frac{X_{jk}}{\pi_k} = t_{X_j}$$

L'idée centrale de l'équilibrage sera alors de chercher un algorithme qui va permettre d'atteindre l'objectif suivant :  $\forall j \in \{1 \dots p\}, V(\hat{t}_{X_j}) = 0$ . Ce sont les contraintes d'équilibrages.

En pratique, le tirage équilibré est implémenté par la méthode du Cube décrite dans Deville et Tillé (2004). De façon synthétique, la méthode est composée de deux phases :

- une première phase dite de « vol » : il s'agit d'une marche aléatoire dans l'espace des contraintes où à chaque étape, l'algorithme arbitre sur l'appartenance à l'échantillon ou non d'une des unités de la base de sondage ;
- une seconde phase dite d'« atterrissage » : lorsque le nombre d'unités sur lesquelles on n'a pas arbitré atteint le nombre de contraintes  $p$  d'équilibrage, alors on relâche la dernière contrainte et on lance une nouvelle phase de vol sur  $p - 1$  contraintes pour arbitrer pour une unité supplémentaire, puis on retire la  $p - 1^{\text{ème}}$  contrainte, etc.

Ainsi l'ordre des variables d'équilibrage a son importance<sup>2</sup> !

## 1.2 Propriétés « pratiques » du tirage équilibré

En pratique, et comme mentionné précédemment, l'équilibrage est rarement exact, et la variance de l'estimation des variables auxiliaires n'est pas nulle... surtout pour les dernières variables d'équilibrage quand elles sont nombreuses. Or, souvent, le tireur peut être tenté de mettre un nombre important de variables d'équilibrage au risque de dégrader la variance sur les dernières variables auxiliaires utilisées.

Une bonne pratique cruciale est de choisir comme première variable d'équilibrage, la variable des probabilités d'inclusion. En effet, ainsi cette contrainte d'équilibrage ne sera jamais relâchée<sup>3</sup> dans la phase d'atterrissage de l'algorithme du cube et c'est ce qui permettra de garantir la taille fixe de l'échantillon :

$$V(\hat{t}_\pi) = 0, i. e. \sum_{k \in S} \frac{\pi_k}{\pi_k} = \sum_{k \in U} 1 = n = Card(S)$$

<sup>1</sup> En pratique, un équilibrage exact est en général impossible, du moins sur l'ensemble des variables d'équilibrage, et l'équilibrage ne sera qu'approché pour un certain nombre de variables (cf. phase d'atterrissage de l'algorithme du cube détaillée ci-après...).

<sup>2</sup> L'équilibrage sera en effet d'autant plus précis (*i.e.* proche d'un équilibrage exact), que les variables seront placées parmi les premières dans la liste des variables d'équilibrage.

<sup>3</sup> Pour peu que la somme des probabilités d'inclusion soit bien entière évidemment, *i.e.* que le plan sondage considéré soit de taille fixe.

Par ailleurs, les variables d'équilibrage ne doivent pas être colinéaires : il faut penser à retirer une variable pour chaque groupe de variables d'équilibrage lorsqu'il y a un risque de colinéarité. C'est par exemple le cas pour les contraintes d'équilibrage issues de variables qualitatives pour lesquelles chaque modalité devient une variable binaire d'équilibrage : il suffit alors de ne pas tenir compte de l'une des modalités en tant que contrainte d'équilibrage.

## 2 Présentation de l'exemple de l'enquête ESS

L'Enquête sociale européenne (ESS) est un vaste dispositif d'enquête comparée et répétée en Europe destiné à la recherche en sciences humaines et sociales. Depuis 2002, il s'agit de mesurer tous les 2 ans le changement des sociétés européennes, à travers des questions d'opinions, d'attitudes et d'évaluation subjective des situations.

### 2.1 Le volet français de l'enquête

Le volet français d'ESS est coordonné par Sciences Po. Il s'agit d'une enquête en face-à-face, dont le champ est celui des individus de 15 ans ou plus vivant en France métropolitaine hors Corse (on note  $N$  la taille de cette population).

Le territoire de la France métropolitaine hors corse est composé de 5036 unités primaires qui correspondent à une agrégation de communes voisines dans un département donné.



Figure 1 : Découpage de la France métropolitaine hors Corse en 5036 unités primaires.

Afin de concentrer la collecte, un premier tirage d'un échantillon de 305 unités primaires a été effectué. La partie exhaustive de cet échantillon d'unités primaires est composée des 45 arrondissements des villes Paris, Lyon et Marseille qui compte chacun pour une unité primaire.

La maîtrise d'ouvrage avait comme exigences de tirer un échantillon à taille fixe, selon des probabilités proportionnelles au nombre d'individus de 15 ans ou plus par UP noté  $N_{up}$ , et sur les variables d'équilibrage suivantes :

- Nombre d'individus de 15 ans ou plus par sexe x tranche d'âge ;
- Nombre d'individus de 15 ans ou plus par décile de niveau de vie ;
- Nombre d'individus de 15 ans ou plus par type de logement ;
- Nombre d'individus de 15 ans ou plus par type de ménage fiscal ;
- Nombre d'individus de 15 ans ou plus par région.

Pour garantir la taille fixe, la variable d'équilibrage des probabilités d'inclusion des UP est

calculée ainsi<sup>4</sup> :  $\pi_{UP} = n \cdot \frac{N_{UP}}{N}$ , avec  $n$  la taille de l'échantillon.

## 2.2 Illustration des propriétés du tirage équilibré

Afin d'illustrer les propriétés évoquées précédemment, des simulations de tirage équilibré de 305 unités primaires ont été réalisées sur trois scénarios d'équilibrage, en mettant dans un premier temps de côté systématiquement les variables de nombre d'individus par région :

- Tirage équilibré avec les probabilités d'inclusion pour seule variable d'équilibrage (équilibrage sur les  $\pi_k$  : scénario de référence) ;
- Tirage équilibré avec toutes les variables d'équilibrage citées et les probabilités d'inclusion en première position (équilibrage classique) ;
- Tirage équilibré avec toutes les variables d'équilibrage citées et les probabilités d'inclusion en dernière position ( $\pi_k$  en dernière position).

Pour évaluer les scénarios, nous disposons de variables de contrôle :

- Nombre d'individus de 15 ans ou plus percevant un revenu salarial ;
- Nombre d'individus de 15 ans ou plus percevant un revenu d'allocations chômage ;
- Nombre d'individus de 15 ans ou plus sous le seuil de pauvreté.

Dans un premier temps, nous regardons la taille des échantillons d'UP tirés pour 10 000 simulations réalisées par scénario :

Scénario	304 UP	305 UP	306 UP
Équilibrage sur les $\pi_k$	0	10 000	0
Équilibrage classique	0	10 000	0
$\pi_k$ en dernière position	1 029	7 916	1 055

Table 1 : Répartition des tailles d'échantillon par scénario (10 000 simulations)

Le scénario dans lequel les probabilités d'inclusion n'est pas la première variable d'équilibrage n'est pas à taille fixe dans environ 20 % des cas. La bonne pratique de mettre les  $\pi_k$  en première variable d'équilibrage n'est pas négociable pour s'assurer un échantillon à taille fixe. Ne pas respecter la taille fixe aura pour conséquence d'augmenter la variance des estimations puisque le nombre d'individus dans le champ de la population ne sera pas correctement estimé, ce qui détériorera la précision sur toutes les autres variables d'équilibrage.

Scénario	T <sub>pop champ</sub>
Équilibrage sur $\pi_k$	0
Équilibrage classique	0
$\pi_k$ en dernière position	0,16

Table 2 : Coefficient de variation de la taille de la population dans le champ

Par ailleurs, l'utilisation de variables auxiliaires permet de gagner nettement en précision sur les variables d'équilibrage et sur les variables de contrôle. Par exemple, ci-dessous les

<sup>4</sup> Le calcul se fait en pratique sur l'ensemble des UP hors les 45 UP exhaustives correspondant aux arrondissements de Paris, Lyon et Marseille ; aucune autre UP ne devient exhaustive lors du calcul des probabilités d'inclusion.

résultats pour trois des variables d'équilibrage :

Scénario	$T_{25-34ans}$	$T_{célibataire}$	$T_{maison}$
Équilibrage sur $\pi_k$	100	100	100
Équilibrage classique	9	12	11
$\pi_k$ en dernière position	16	16	13

Table 3 : Coefficients de variation avec scénario d'équilibrage avec les  $\pi_k$  en base 100

Nous avons le même constat sur des variables de contrôle :

Scénario	$T_{salariés}$	$T_{chômeurs}$	$T_{sous\ le\ seuil\ pauvreté}$
Équilibrage sur $\pi_k$	100	100	100
Équilibrage classique	33	65	14
$\pi_k$ en dernière position	40	68	15

Table 4 : Coefficients de variation avec scénario d'équilibrage avec les  $\pi_k$  en base 100

L'utilisation de variables d'équilibrage permet également d'améliorer la précision sur des variables n'ayant pas servi à l'équilibrage (variables de contrôle), mais dans une moindre mesure.

On remarque que l'absence de taille fixe ( $\pi_k$  en dernière position) dégrade les estimations à variables d'équilibrage données (équilibrage classique *versus* équilibrage avec les  $\pi_k$  en dernière position).

### 2.3 Intégrer une contrainte spatiale dans les tirages ?

Lorsqu'un indicateur étudié a une forte autocorrélation spatiale positive, collecter par exemple de l'information sur une commune voisine d'une commune déjà tirée apporte peu d'information nouvelle.

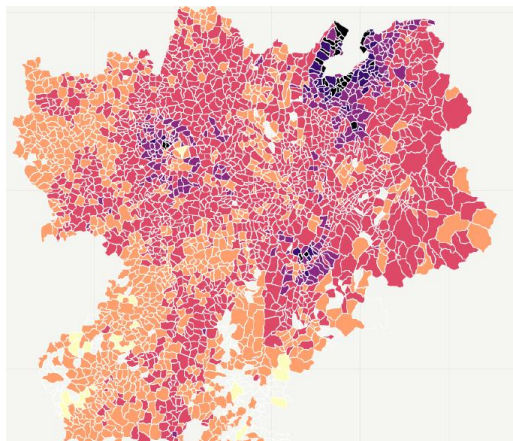


Figure 2 : Niveau de vie communal en Auvergne-Rhône-Alpes

Du coup, tirer des unités bien réparties spatialement permet de gagner en précision sur les variables avec autocorrélation spatiale positive. La méthode du cube locale pour conserver les propriétés du tirage équilibré tout en assurant une bonne répartition spatiale de l'échantillon est décrite dans Grafström et Tillé (2013). Cet algorithme permet d'obtenir un échantillon qui

est à la fois étalé géographiquement et équilibré sur des variables auxiliaires.

Sur notre exemple de l'enquête ESS, même sans utiliser les variables d'équilibrage, il y a un gain en précision en ayant recours à un tirage spatialement équilibré, par rapport à un équilibrage uniquement sur les  $\pi_k$ .

Scénario	T <sub>25-34ans</sub>	T <sub>célibataire</sub>	T <sub>maison</sub>
Équilibrage sur $\pi_k$	100	100	100
Équilibrage spatial sur $\pi_k$	76	94	69
Équilibrage spatial classique	10	14	12
Pour mémoire : équilibrage classique	9	12	11

Table 5 : Coefficients de variation avec scénario d'équilibrage avec les  $\pi_k$  en base 100

## 2.4 Comparaison entre les deux tirages

Dans l'exemple de l'enquête ESS, sur les variables d'équilibrage, l'utilisation d'un tirage spatialement équilibré ne dégrade que très légèrement la précision par rapport à un tirage simplement équilibré (cf. Table 5).

Cette dégradation est à la fois logique, puisque des contraintes ont été ajoutées, et rassurante vu la faible hausse des coefficients de variation.

Sur les variables de contrôle, celles sur lesquelles il n'y a pas de contrainte d'équilibrage, il y a une nette amélioration de la précision avec un équilibrage spatial.

Scénario	T <sub>salariés</sub>	T <sub>chômeurs</sub>	T <sub>sous le seuil pauvreté</sub>
Équilibrage sur $\pi_k$	100	100	100
Équilibrage spatial sur $\pi_k$	66	84	82
Équilibrage classique	33	65	14
Équilibrage spatial classique	25	51	14

Table 6 : Coefficients de variation avec scénario d'équilibrage avec les  $\pi_k$  en base 100

L'absence d'amélioration de la précision pour le nombre de personnes sous le seuil de pauvreté vient du fait que cette variable s'explique en grande partie par la variable d'équilibrage de nombre d'individus dans le 1<sup>er</sup> décile de niveau de vie.

Si l'on maintient la précision sur les variables d'équilibrage, on l'améliore sur les variables de contrôle, mais *quid* de variables avec une structure géographique marquée ?

Dans l'exemple de l'enquête ESS, sur les variables de totaux d'individus par région<sup>5</sup>, le tableau 7 montre que l'équilibrage classique ne permet pas de gagner (ou très peu) en précision dans les régions les moins peuplées (Picardie et Bretagne dans l'exemple).

---

<sup>5</sup> À ce stade, nous avons écarté les totaux d'individus régionaux des variables d'équilibrage. Ce sont donc pour le moment des variables de contrôle.

Scénario	$T_{IDF}$	$T_{Picardie}$	$T_{Nord PdC}$	$T_{Bretagne}$
Équilibrage sur $\pi_k$	100	100	100	100
Équilibrage classique	57	98	74	91
Équilibrage spatial sur $\pi_k$	22	47	21	25
Équilibrage spatial classique	27	76	41	59

Table 7 : Coefficients de variation avec scénario d'équilibrage avec les  $\pi_k$  en base 100

En revanche, l'équilibrage spatial sur les  $\pi_k$  permet de gagner nettement en précision sur la taille de population dans toutes les régions. Toutefois quand le nombre de variables augmente dans l'équilibrage spatial (de 1 lorsque seuls les  $\pi_k$  sont utilisés à 30, ici pour l'équilibrage spatial classique), on perd nettement en précision sur la taille de population dans toutes les régions. En résumé, à cet instant, on gagne en précision sur les variables nationales en augmentant le nombre de variables d'équilibrage, mais on perd en précision régionale.

### 3 Comment améliorer la précision régionale et nationale ?

Peut-on gagner sur tous les plans quand on augmente le nombre de variables d'équilibrage dans un équilibrage spatial ? *i.e.* peut-on améliorer la précision sur les variables locales (totaux régionaux, ici) et sur l'ensemble des variables nationales ?

#### 3.1 Une fausse bonne idée : glisser les variables régionales parmi les variables d'équilibrage

Pour évaluer cette piste, nous allons évaluer deux scénarios supplémentaires, de tirages équilibrés, spatialement ou non, qui intègrent comme contraintes d'équilibrage, les totaux d'individus par région.

Scénario	$T_{25-34ans}$	$T_{célibataire}$	$T_{maison}$
Équilibrage sur $\pi_k$	100	100	100
Équilibrage classique	9	12	11
Équilibrage spatial classique	10	14	12
Équilibrage classique avec variables régionales	9	12	11
Équilibrage spatial avec variables régionales	22	20	15

Table 8 : Coefficients de variation avec scénario d'équilibrage avec les  $\pi_k$  en base 100

En ajoutant les variables régionales d'équilibrage, il n'y a pas de dégradation de la précision du tirage équilibré classique, mais la précision se dégrade dans le cadre d'un tirage spatialement équilibré. Ce constat sur les variables d'équilibrage se retrouve également sur les variables de contrôle.

La dégradation de la précision dans le cadre d'un tirage spatialement équilibré s'explique par la perte de la taille fixe, alors même que les probabilités d'inclusion sont en première position de l'équilibrage.

Scénario / UP	303	304	305	306	307	308	309	310
Équilibrage classique avec variables régionales	0	0	10 000	0	0	0	0	0
Équilibrage spatial avec variables régionales	48	1837	6243	1823	45	3	0	1

Table 9 : Répartition des tailles d'échantillon par scénario (10 000 simulations)

En revanche, la précision s'améliore, logiquement, assez nettement sur les variables régionales :

Scénario	$T_{IDF}$	$T_{Picardie}$	$T_{Nord PdC}$	$T_{Bretagne}$
Équilibrage sur $\pi_k$	100	100	100	100
Équilibrage classique	57	98	74	91
Équilibrage spatial classique	27	76	41	59
Équilibrage classique avec variables régionales	14	19	15	18
Équilibrage spatial avec variables régionales	13	18	18	23

Table 10 : Coefficients de variation avec scénario d'équilibrage avec les  $\pi_k$  en base 100

Le problème du non-respect de la taille fixe est probablement lié à une incompatibilité entre l'algorithme du cube local (arbitrage de sélection d'UP entre des unités voisines) et des contraintes d'équilibrage locales.

Le même problème a été rencontré lors de l'étude préalable au tirage de l'échantillon-maître Nautile avec des variables de TUU<sup>6</sup>, problème également retrouvé lors de l'étude de tirage d'unités primaires pour le dispositif de biosurveillance Albane de Santé publique France.

## 3.2 Deux autres solutions...

Deux autres solutions sont envisageables :

- Remplacer les variables d'équilibrage, y compris donc les variables avec une structure géographique, par les premiers axes d'une ACP ;
- Utiliser un tirage spatialement équilibré stratifié.

### 3.2.1 Utilisation d'axes d'ACP

Cette solution consiste à remplacer les variables d'équilibrage par des axes d'ACP, tout en conservant à part la variable de probabilités d'inclusion en première position dans l'équilibrage.

Lors de la mise en œuvre de cette solution, il est possible de conserver dans l'équilibrage des variables brutes (*i.e.* non intégrées dans l'ACP) et des variables d'axes d'ACP (Chevalier et al. (2022) au sujet de l'échantillon-maître Nautile), à condition de disjointer les variables utilisées de manière brute dans l'équilibrage et celles utilisées dans l'ACP.

<sup>6</sup> Tranches d'unité urbaine

Une question porte naturellement sur le nombre d'axes d'ACP à conserver, puisqu'il doit permettre de conserver la taille fixe de l'échantillon afin de ne pas dégrader la précision, tout en expliquant le maximum de variance de l'ACP. Empiriquement, dans les exemples étudiés côté enquêtes ménage, conserver les axes qui expliquent 99 %, voire 99,9 %, de la variance de l'ACP permet de conserver la taille fixe. Dans notre exemple de l'enquête ESS, l'ACP porte sur 49 variables d'équilibrage, et pour conserver 99,9 % de la variance, 29 axes ont été retenus.

Scénario	T <sub>25-34ans</sub>	T <sub>célibataire</sub>	T <sub>maison</sub>	T <sub>salariés</sub>	T <sub>chômeurs</sub>	T <sub>sous le seuil pauvreté</sub>
Équilibrage sur $\pi_k$	100	100	100	100	100	100
Équilibrage classique	9	12	11	33	65	14
Équilibrage spatial classique	10	14	12	25	51	14
Équilibrage classique avec variables régionales	9	12	11	30	57	13
Équilibrage avec ACP	22	23	8	34	60	16
Équilibrage spatial avec ACP	15	16	8	25	49	14

Table 11 : Coefficients de variation avec scénario d'équilibrage avec les  $\pi_k$  en base 100

On observe plutôt une bonne performance sur les variables d'équilibrage (3 premières variables du tableau 11) mais on perd tout de même de la précision par rapport à un tirage équilibré sur les variables en brut, surtout pour les premières variables de l'équilibrage initial. Ceci s'explique logiquement par le fait que des axes de l'ACP non retenus pour effectuer l'équilibrage expliquent partiellement des variables auparavant intégrées en brut dans l'équilibrage. Sur les variables de contrôle (3 dernières variables du tableau 11) qui ne participent pas à l'équilibrage, ni en brut ni en tant que composantes des axes de l'ACP, on peut observer un maintien du niveau de précision.

Scénario	T <sub>IDF</sub>	T <sub>Picardie</sub>	T <sub>Nord PdC</sub>	T <sub>Bretagne</sub>
Équilibrage sur $\pi_k$	100	100	100	100
Équilibrage classique	57	98	74	91
Équilibrage spatial classique	27	76	41	59
Équilibrage classique avec variables régionales	14	19	15	18
Équilibrage avec ACP	11	12	8	14
Équilibrage spatial avec ACP	9	12	7	14

Table 12 : Coefficients de variation avec scénario d'équilibrage avec les  $\pi_k$  en base 100

Par contre, sur les totaux régionaux, il y a un gain assez net à utiliser un équilibrage spatial sur des axes d'ACP intégrant des contraintes régionales, par rapport par exemple à un équilibrage spatial classique avec des variables d'équilibrage brutes sans contraintes régionales.

### 3.2.2 Équilibrage (spatial) stratifié

Pour réaliser un équilibrage stratifié, on utilise l'algorithme de tirage doublement équilibré – spatialement dispersé et équilibré sur des variables auxiliaires – stratifié avec mise en commun des phases d'atterrissage implémenté dans le package *BalancedSampling* (fonction *lcubestratified*) et basé sur Chauvet (2009) et Grafström et Tillé (2013). Dans ce contexte, on remplace les probabilités d'inclusion nationales par des probabilités d'inclusions par strate, ici régionales dans le cas de notre exemple de l'enquête ESS :  $\pi_{UP} = n_{région} \cdot \frac{N_{UP}}{N_{région}}$

Avec des scénarios d'équilibrage stratifié, on observe une légère dégradation de la précision sur les variables nationales (contraintes d'équilibrages ou variables de contrôle), potentiellement due à la déformation des probabilités d'inclusion et au fait que la version équilibrée stratifiée rajoute des contraintes de tailles fixes par strate à l'équilibrage, ce qui dégrade forcément la précision de l'équilibrage au niveau national par rapport aux versions non stratifiées.

Scénario	T <sub>25-34ans</sub>	T <sub>célibataire</sub>	T <sub>maison</sub>	T <sub>salariés</sub>	T <sub>chômeurs</sub>	T <sub>sous le seuil pauvreté</sub>
Équilibrage sur $\pi_k$	100	100	100	100	100	100
Équilibrage classique	9	12	11	33	65	14
Équilibrage spatial classique	10	14	12	25	51	14
Équilibrage classique avec variables régionales	9	12	11	30	57	13
Équilibrage avec ACP	22	23	8	34	60	16
Équilibrage spatial avec ACP	15	16	8	25	49	14
Équilibrage stratifié	20	23	17	30	54	20
Équilibrage spatial stratifié	22	25	18	29	52	23

Table 13 : Coefficients de variation avec scénario d'équilibrage avec les  $\pi_k$  en base 100

En revanche, on enregistre une très nette amélioration de la précision sur les variables de taille du champ dans chaque région puisque la précision est désormais parfaite.

Scénario	T <sub>IDF</sub>	T <sub>Picardie</sub>	T <sub>Nord PdC</sub>	T <sub>Bretagne</sub>
Équilibrage sur $\pi_k$	100	100	100	100
Équilibrage classique	57	98	74	91
Équilibrage spatial classique	27	76	41	59
Équilibrage classique avec variables régionales	14	19	15	18
Équilibrage avec ACP	11	12	8	14
Équilibrage spatial avec ACP	9	12	7	14
Équilibrage stratifié	0	0	0	0
Équilibrage spatial stratifié	0	0	0	0

Table 14 : Coefficients de variation avec scénario d'équilibrage avec les  $\pi_k$  en base 100

Nous aurions aussi pu construire des variables de contrôle régionales pour pouvoir vérifier que ces tirages équilibrés stratifiés et spatialement équilibrés stratifiés améliorent aussi la précision régionale d'autres variables.

Au final, pour l'enquête ESS, le choix d'un tirage spatialement équilibré stratifié a été fait. Ce choix peut sembler surprenant au regard des tableaux présentés mais au-delà des phénomènes d'amélioration ou de dégradation observés relativement à un scénario de référence, il faut regarder les coefficients de variation en « niveau ».

Scénario	T <sub>25-34ans</sub>	T <sub>célibataire</sub>	T <sub>salariés</sub>	T <sub>chômeurs</sub>	T <sub>IDF</sub>	T <sub>Picardie</sub>
Équilibrage sur $\pi_k$	1,16	1,57	0,63	0,87	10,42	32,35
Équilibrage classique	0,10	0,19	0,21	0,57	5,97	31,75
Équilibrage spatial classique	0,11	0,21	0,16	0,44	2,77	24,46
Équilibrage classique avec variables régionales	0,10	0,19	0,19	0,50	1,48	6,28
Équilibrage avec ACP	0,25	0,37	0,21	0,52	1,18	3,74
Équilibrage spatial avec ACP	0,17	0,25	0,16	0,43	0,95	3,74
Équilibrage stratifié	0,24	0,36	0,19	0,46	0,0	0,0
Équilibrage spatial stratifié	0,26	0,39	0,18	0,45	0,0	0,0

Table 15 : Coefficients de variation en niveau

Du coup, au regard des faibles niveaux de coefficients de variation pour les variables nationales (4 premières variables du tableau 15), même si une légère dégradation est observée, cela ne pose pas de problème. Par contre, pour les variables régionales (2 dernières colonnes du tableau 15), les coefficients de variation étaient assez élevés, et pour le coup les réduire nettement était important au regard d'objectifs de précision régionale.

## 4 Conclusion

Avant de résumer les points clés développés lors de cette étude, il convient de mentionner que les simulations ont été réalisées au moyen d'outils *R* internes développés par Martin Chevalier (Insee) et utilisés pour tous les tirages d'unités primaires depuis 2018.

L'objectif de ces outils est de permettre au sondeur de se concentrer uniquement sur les aspects méthodologiques de l'étude en réduisant les préoccupations informatiques liées à la multitude de simulations.

En pratique, le sondeur utilise les fonctions développées pour définir le ou les scénarios de tirage après avoir construit la base de sondage des unités primaires. Les simulations seront alors générées pour chacun des scénarios. Puis, une fonction de synthèse permet de compiler automatiquement toutes les simulations par scénario afin d'obtenir les biais relatifs, les CV, etc.

Au final, les simulations montrent que l'équilibrage spatial améliore la précision sur les variables de l'enquête ayant une autocorrélation spatiale positive en dégradant très peu la qualité globale l'équilibrage.

Des précautions doivent être prises lorsqu'on intègre des variables d'équilibrage locales (régionale, départementale, par TUU...).

Deux solutions existent alors pour améliorer la précision sur des variables locales : il est en effet possible de réaliser préalablement une ACP puis d'effectuer un équilibrage sur les « premiers axes » d'ACP en étant vigilant au nombre d'axes retenus, ou encore d'effectuer un tirage spatialement équilibré stratifié.

Ces deux approches ne sont pas mutuellement exclusives mais peuvent au contraire se combiner pour assurer à la fois une stratification régionale et une prise en compte de zonages du type TUU.

## Bibliographie

Costa, L. et Merly-Alpa, T. (2017), *L'échantillonnage équilibré*, Notes méthodologiques de l'Insee.

Chauvet, G. (2009), *Stratified balanced sampling*, Survey Methodology, 35, 115-119.

Deville, J.-C. et Tillé, Y. (2004), *Efficient Balanced Sampling : The Cube Method*, Biometrika, vol. 91, No 4, pp. 893-912.

Deville, J.-C. et Tillé, Y. (2005), *Variance approximation under balanced sampling*, Journal of Statistical Planning and Inference 128, pp 569-591.

Grafström, A. et Tillé, Y. (2013), *Doubly balanced spatial sampling with spreading and restitution of auxiliary totals*, Environmetrics 24.2, pp 120–131.