

---

## Estimation de la variance liée au nouveau plan de sondage de l'Enquête Emploi en continu (EEC)

*Lionel DELTA (\*), Nicolas PALIOD (\*)*

*(\*) Insee, Direction de la méthodologie et de la coordination statistique et internationale  
Division Sondages*

lionel.delta@insee.fr , nicolas.palioid@insee.fr

**Mots-clés.** : Estimation de variance, échantillons équilibrés, sondage indirect, échantillonnage à deux degrés

**Domaines.** : Calcul de précision, estimation de variance.

---

### Résumé

**Résumé.** L'Institut national de la statistique et des études économiques (Insee) a réformé à partir de 2019 et pour la décennie à suivre, le plan de sondage de l'Enquête emploi en continu (EEC) en France métropolitaine.

Une des particularités de l'EEC est qu'il s'agit d'une enquête aréolaire, c'est-à-dire que son échantillon est un ensemble de zones géographiques compactes appelées grappes, regroupées en secteurs comprenant six ou sept grappes. Le tirage des secteurs se fait au sein d'unités de coordination qui sont tirées lors d'une première phase de tirage ; le tirage des secteurs, puis celui des grappes constituent alors les phases suivantes.

Comme chacune de ces phases ajoute de l'aléa dans les estimations, le calcul de variance dans l'enquête Emploi nécessite de savoir exprimer la variance générée par chaque phase de tirage puis de savoir articuler l'ensemble de ces contributions pour en déduire la variance totale d'estimation.

Par rapport au plan de sondage précédent, le recours à des unités de coordination est complètement nouveau conduisant à un tirage des secteurs, non plus en une seule étape mais en plusieurs phases. En outre, le tirage des unités de coordination se fait de façon indirecte à partir de l'échantillon-maître - l'échantillon de premier degré - des autres enquêtes ménages. De ce fait il ne s'agit pas d'un plan de sondage classique auquel on peut spécifiquement associer des formules d'estimation de variance décrites dans la littérature.

Des études méthodologiques ont permis d'entériner le principe d'estimation de la variance de tirage des secteurs selon une formule de décomposition de la variance valable pour les plans de sondage à plusieurs degrés avec une utilisation particulière de l'estimateur de Sen-Yates-Grundy en première phase pour prendre en compte le tirage unités de coordination.

# Abstract

The French National Institute of Statistics and Economics Studies recently reformed, for the next decade, the sample design of the French Labour Force Survey. Its sample is a set of compact geographic areas : sectors are drawn from coordinating units that are drawn in a first phase of sampling ; clusters are then selected from the previously drawn sectors.

Since each of these phases adds randomness to the estimates, the calculation of variance in the Labour Force Survey requires knowing how to express the variance generated by each phase of drawing and then knowing how to articulate all of these contributions in order to deduce the total estimation variance.

Methodological studies have made it possible to endorse the principle of estimating the variance of the sectors drawn according to a formula for decomposing the variance that is valid for multi-stage sampling designs, with particular use of the Sen-Yates-Grundy estimator in the first stage.

## Introduction

L'enquête Emploi apporte de nombreuses informations sur l'état du marché du travail et son évolution : le volume de la population active, le nombre de chômeurs et le taux de chômage, les caractéristiques des personnes présentes sur le marché du travail, de celles qui arrivent ou le quittent, etc.

L'échantillon Emploi est constitué, pour la France hors régions et territoires d'Outre-mer, d'environ 70 000 logements identifiés comme résidence principale. L'échantillon comprend également un certain nombre de logements non principaux et de logements considérés comme vacants. L'EEC se caractérise par une stratégie d'échantillonnage qualifiée d'aérolaire : l'échantillon de cette enquête est constitué, au niveau le plus fin, par un tirage de groupes de logements contigus appelés grappes.

Le tirage de la première phase de l'EEC a été coordonnée avec celui des autres enquêtes ménages pour les années à venir, par le biais de l'Echantillon-Maitre (EM).

Cette coordination dans le tirage offrait la possibilité d'envisager une méthodologie commune à l'ensemble des enquêtes ménages pour l'estimation de variance de première phase. De ce fait les travaux menés ont permis d'adopter une méthodologie pour rendre compte de la variance générée par le tirage de l'EM. Celle-ci se base sur la formule de variance de Deville-Tillé et devait théoriquement permettre le calcul de variance de première phase dans l'EEC à partir des unités primaires de l'EM.

Cependant, l'adaptation de cette méthodologie à l'EEC conduisait à des estimations nettement biaisées, particulièrement au niveau régional.

À défaut de pouvoir mutualiser efficacement le calcul de variance de première phase des différentes enquêtes, les travaux méthodologiques ont été recentrés sur les unités de coordination de l'EEC en recourant à l'estimateur de variance de Sen-Yates-Grundy.

## 1 Contexte : la refonte du plan de sondage de l'enquête Emploi à l'Insee

Pour bien mesurer les évolutions trimestrielles d'emploi et de chômage, une part importante de l'échantillon est conservée d'un trimestre sur l'autre : seul 1/6ème de l'échantillon est renouvelé chaque trimestre. Ainsi, chaque logement de l'échantillon est interrogé pendant 6 trimestres

consécutifs.

Du 3ème trimestre 2019 au 4ème trimestre 2020, un nouveau plan de sondage est progressivement monté en charge en remplacement de la méthodologie mise en place en 2009.

Les logements de la base de sondage ont ainsi été regroupés par unités géographiques à différents niveaux et qui sont autant d'unités statistiques pertinentes pour l'Emploi :

- La grappe Emploi : elle est composée approximativement de 20 résidences principales (RP) au moment de sa constitution. S'y ajoutent ensuite des logements vacants, des résidences non principales et des logements neufs afin d'avoir une représentativité de logements nouvellement devenus des résidences principales au moment de la collecte. On dénombre en France métropolitaine un total de 1 395 866 grappes.
- Le secteur Emploi : il correspond à un regroupement de 6 (parfois 7) grappes n'appartenant pas forcément à la même commune ni au même IRIS<sup>1</sup>. Les grappes d'un secteur donné appartiennent à la même unité de coordination (cf. définition ci-après). Il y a 231 966 secteurs.
- L'unité primaire (UP) : composée d'une ou plusieurs communes avec la contrainte de taille minimale de 2500 résidences principales. Le territoire métropolitain a été découpé en 5064 unités primaires appartenant chacune à une unique ancienne région.
- L'unité de coordination (UC) : au nombre de 1646, elles correspondent à un regroupement de plusieurs dizaines voire plusieurs centaines de secteurs emploi (en moyenne 140 secteurs). Certaines d'entre elles (245 UC) coïncident avec une unité primaire, les autres (1401 UC) forment un groupe de 2 à 7 UP.

À l'instar de ce qui était déjà en place lors de la décennie précédente, l'échantillon de secteurs est commun à tous les échantillons trimestriels tirés pendant plusieurs années jusqu'à ce que celui-ci soit remplacé par un nouvel échantillon.

Une fois l'échantillon fixe de secteurs constitué<sup>2</sup>, le plan de sondage des logements au sein de ces secteurs n'a pas été fondamentalement modifié à l'occasion de la refonte. Par conséquent cette partie n'entraîne pas de modifications méthodologiques majeures pour estimer la variance.

Cette étude se concentre donc sur l'estimation de la variance d'échantillonnage des secteurs selon les étapes suivantes :

1. Tirage des UP : Les UP sont tirées suivant un plan de sondage, stratifié selon les anciennes régions administratives françaises, doublement équilibré avec étalement spatial et atterrissage national. L'échantillon d'UP constitue ainsi le premier degré de tirage pour les enquêtes ménages habituellement réalisées à l'Insee, à l'exception de l'enquête Emploi pour des raisons de taille d'échantillon.
2. Tirage indirect des UC : l'échantillon d'UC est constitué par identification des UC contenant au moins une UP tirée à l'étape précédente. Il s'agit donc d'un plan de sondage indirect pour ces unités de coordination dont les poids sont calculés à partir des poids des UP suivant la méthode généralisée du partage des poids dans sa version pondérée.
3. Tirage des secteurs : les secteurs sont tirés suivant un plan de sondage doublement équilibré avec étalement spatial, stratifié par UC avec atterrissage régional. Les probabilités

---

1. Le sigle IRIS - îlots regroupés pour l'information statistique - fait référence à un découpage infra-communal réalisé par l'INSEE. En pratique, l'IRIS correspond à l'échelon géographique le plus fin pour lequel des statistiques, essentiellement démographiques, sont diffusées en France.

2. Lors de la décennie précédente, l'échantillon de première phase avait été sélectionné directement par tirage équilibré de secteurs Emploi au sein de chaque région, soit en une seule étape.

d'inclusion sont égales pour tous les secteurs appartenant à une même UC.

## 2 Recherche d'un estimateur de variance spécifique au plan de sondage de l'EEC : méthodologie et données utilisées

Les travaux de recherche d'une formule permettant de mesurer, avec de bonnes propriétés statistiques, la variance des estimateurs construits à partir des échantillons de l'EEC ont été basés sur une adaptation d'estimateurs déjà proposés dans la littérature scientifique.

Le choix des estimateurs de variance retenus ainsi que les adaptations faites ont été validées empiriquement en confrontant la distribution de ces estimations de variance portant sur des jeux de données simulés et les niveaux de variance attendus sous le plan de sondage. Ces niveaux de variance-cibles n'étant pas exactement connus, d'où le besoin de les estimer, ils sont approchés pour ces variables de validation par les variances empiriques obtenues par simulation.

Dans la pratique, plusieurs bases de données ont été construites à partir de différentes sources (par exemple les données fiscales mais aussi le recensement de la population) dans le but de valider la stratégie d'estimation de la variance de l'enquête Emploi. Ces bases contiennent des variables qui jouent le rôle des variables d'intérêt de l'enquête et qui sont donc connues ici pour l'ensemble des unités statistiques pertinentes (UC, UP, secteurs-Emploi, etc).

On peut les classer selon deux grandes familles de variables. Le premier type correspond aux variables construites à partir de la source fiscale qui a servi à constituer les variables d'équilibrage du plan de sondage des UP et des secteurs. La précision de l'estimation du total de ces variables est du même ordre que celle des variables d'équilibrage. De fait, elles ont une très bonne précision et correspondent quasiment à des variables d'équilibrages.

La seconde famille de variables est essentiellement issue du RP. Ces variables pourraient mieux refléter les propriétés des variables d'enquête : il y a par construction des corrélations avec les variables d'équilibrage mais de façon plus limitée et ces variables ne sont donc pas directement équilibrées comme c'est le cas des premières variables.

Le plan de sondage de l'enquête Emploi est répété un très grand nombre de fois permettant d'obtenir plusieurs milliers d'échantillons dits simulés (un échantillon simulé correspondant à l'ensemble des UC, secteurs et grappes tirés lors d'une simulation du processus de tirage).

À partir des échantillons simulés, des indicateurs nationaux et régionaux analogues à ceux mesurés dans l'enquête Emploi (effectifs dans les différents statuts d'activité par âge et par sexe) ainsi que des indicateurs portant sur d'autres thématiques peuvent être mesurés ainsi que la variance empirique de ces indicateurs sur l'ensemble des simulations, que l'on assimile à la variance d'échantillonnage.

Pour chaque indicateur calculé, on cherche donc à estimer cette variance d'échantillonnage (connue empiriquement) à partir d'expressions concurrentes d'estimateurs de variance.

L'appréciation des propriétés statistiques des différentes formules de variances testées se fait à partir des mesures d'indicateurs de qualité empirique sur l'ensemble des simulations : le biais empirique, la variance empirique et le taux de couverture de l'intervalle de confiance à 95%.

### 3 Estimation de la variance liée au tirage des secteurs Emploi

#### 3.1 Méthode d'estimation de la variance pour un plan de sondage à deux degrés, la formule de Rao

L'estimation de variance de l'échantillon de logements de l'EEC repose, comme pour toute enquête, sur la capacité à rendre compte de la variance générée par chacune des étapes ayant conduit à l'échantillon final. La variance totale correspond alors à une articulation de la variance de ces différentes étapes à travers une formule adaptée au type de plan de sondage.

La plupart des échantillons des enquêtes ménages à l'Insee sont obtenus à partir de plans de sondage à plusieurs degrés. Pour de tels plans de sondage, le calcul de variance est basé sur l'utilisation d'une formule très utilisée à l'Insee, appelée formule de Rao (1975).

Pour peu que l'on dispose :

- d'un estimateur sur l'échantillon des unités primaires de la variance d'un estimateur de total d'une variable  $y$  associée au premier degré de sondage qui soit sans biais et sous forme quadratique

$$\hat{V}_{UP}(y_u, u \in s_{UP}) = Q(y_u, u \in s_{UP}) = \sum_{u \in s_{UP}} q_u y_u^2 + \sum_{(u,v) \in s_{UP}, u \neq v} q_{uv} y_u y_v$$

avec  $y_u$  la valeur de la variable  $y$  pour l'unité primaire  $u$  appartenant à l'échantillon d'unités primaires  $s_{UP}$  et les  $q_u$  et  $q_{uv}$  les coefficients de la forme quadratique

- d'une série d'estimateurs, sur l'échantillon  $s_u$  des unités secondaires  $z$  tirées au sein de chaque unité primaire  $u$ , de la variance de l'estimateur du total sur  $u$  associée au second degré de sondage  $\hat{V}_{US,u}(y_z, z \in s_u)$ , qui soient sans biais

alors l'estimateur suivant estime sans biais et sur l'échantillon des unités secondaires la variance associée aux deux degrés de sondage :

$$\hat{V}(y_z, z \in s_{US}) = \hat{V}_{UP}(\hat{y}_u, u \in s_{UP}) + \sum_{u \in s_{UP}} \left( \frac{1}{\pi_u^2} - q_u \right) \hat{V}_{US,u}(y_z, z \in s_u) \quad (1)$$

avec  $\hat{y}_u = \sum_{z \in s_u} \frac{y_z}{\pi_{z|u}}$ ,  $\pi_u$  la probabilité de sélection de l'unité primaire  $u$  et  $\pi_{z|u}$  la probabilité de sélection de l'unité secondaire  $z$  au sein de l'unité primaire  $u$ .

Quand la forme quadratique de premier degré est un estimateur faiblement biaisé, en pratique l'application de la formule de Rao ne se traduit pas par une amplification du biais et donc l'estimateur sur les unités secondaires est lui-même faiblement biaisé.

Le choix d'une formule de variance rendant compte du tirage des secteurs s'est donc articulée autour de la validation des hypothèses précitées déclinées en 3 volets : la recherche d'un estimateur de variance optimal associé au tirage des unités de coordination, la recherche d'estimateurs de variances conditionnelles associées au tirage des secteurs au sein des UC, l'appréciation de la pertinence de l'hypothèse d'indépendance des tirages des secteurs, hypothèse cruciale pour l'utilisation de la formule de Rao.

#### 3.2 Estimation de la variance liée au tirage des UC

Le plan de sondage des UC décrit en section 1.2 correspond à un plan de sondage indirect pour lequel les caractéristiques habituelles (probabilités de sélection des différentes unités, taille

de l'échantillon, etc) ne sont pas données a priori.

Cependant, l'échantillon d'UC et les pondérations associées se déduisent de l'échantillon d'UP tirés, il est également possible d'exprimer la variance calculée via les UC à partir d'une expression de variance calculée via les UP, en ayant recours à des variables transformées. Dans cette optique, un estimateur de variance rendant compte du tirage des UP peut être adapté pour rendre compte du tirage des UC.

Le plan de sondage des unités primaires étant un plan de sondage équilibré avec étalement spatial, stratifié par région et avec atterrissage national, l'estimateur proposé par Deville et Tillé (1992) apparaît comme un bon candidat. En tenant compte de l'étalement spatial en plus de l'équilibrage, on peut lui préférer l'estimateur de Grafström-Tillé (2013).

Néanmoins, ces estimateurs n'ont de bonnes propriétés statistiques (notamment le caractère faiblement biaisé) qu'à très grande échelle. Autrement dit, le plan de sondage doit être tel que le nombre de contraintes d'équilibrages appliquées soit négligeable devant la taille des échantillons tirés.

Cette hypothèse n'est en pratique pas vérifiée dans le cas du plan de sondage de l'EEC. En effet, le plan de sondage vise à ce qu'au sein de chaque région, l'échantillon tiré soit équilibré sur 56 variables auxiliaires. Le nombre total de contraintes, auquel s'ajoute celle de taille fixe de l'échantillon, s'avère alors bien supérieur au nombre d'UP tirées dans chacune des régions. L'allocation moyenne régionale est d'environ 25 unités primaires mais, dans certaines régions, l'allocation est inférieure à 10. Par conséquent la sélection des échantillons se fait nécessairement par un relâchement de certaines contraintes d'équilibrage, ce qui rend alors les estimateurs de Deville-Tillé et Grafström-Tillé plus difficilement opérants.

Ce relâchement se fait, au moins dans un premier temps, en passant de l'échelle régionale à l'échelle nationale pour le respect des contraintes d'équilibrage de certaines variables. Ainsi, on observe de façon empirique un quasi-équilibre pour deux tiers des variables auxiliaires au niveau national<sup>3</sup>. En revanche, à l'échelle régionale, les variables auxiliaires montrent un écart plus important à la situation d'équilibrage parfait<sup>4</sup>.

Face aux limites de cette première classe d'estimateurs, on considère une seconde classe d'estimateurs non spécifiques aux tirages équilibrés. Il s'agit des estimateurs de variance d'Horvitz-Thompson et de Sen-Yates-Grundy (1953). Ces derniers sont valables pour des plans de sondage de taille fixe et calculables à la condition de disposer, en plus des probabilités d'inclusion simples, des probabilités d'inclusion conjointe de tous les couples d'unités appartenant à l'échantillon. Ces dernières probabilités d'inclusion dites doubles ne sont cependant pas connues a priori pour le plan de sondage des UP et sont obtenues à l'aide de plusieurs millions de simulations de tirages d'échantillons.

Après comparaison à partir des variables d'intérêt construites au niveau UP, les premières analyses empiriques font pencher en faveur de la première classe d'estimateurs, à condition de prendre en compte un nombre réduit de variables d'équilibrage. Les estimateurs de variance correspondant sont alors moins biaisés et ont une variance plus faible que ceux de la seconde classe. Cette première conclusion est contredite par les résultats empiriques sur des variables transformées de niveau UP ; c'est-à-dire des variables initialement construites au niveau UC, et donc celles qui sont pertinentes pour l'EEC, que l'on "redistribue" au niveau UP en fonction des liens pondérés entre ces deux niveaux d'unités.

---

3. Les coefficients de variation des estimateurs Horvitz-Thompson relatifs à ces variables sont inférieurs à 0.5%, voire même à 0.3% pour la plupart d'entre elles. Pour les autres variables auxiliaires, les CV considérés sont souvent supérieurs à 1%

4. En considérant l'ensemble des variables régionalisées, c'est-à-dire issues des croisements entre variables auxiliaires et indicatrices d'appartenance à une région donnée, moins d'une variable d'équilibrage régionale sur dix affiche un CV inférieur à 1% pour l'estimateur Horvitz-Thompson du total.

Dans ce cas, les estimateurs de variance de Deville-Tillé calculés sur ces variables sont beaucoup moins performants ; l'expression de Deville-Tillé aboutit ainsi à des résultats beaucoup plus biaisés que l'estimateur de Sen-Yates-Grundy notamment lorsque ce dernier est calculé directement à partir de l'échantillon d'UC. Calculé au niveau UP, ce dernier estimateur est en revanche plus biaisé et plus instable.

C'est donc l'estimateur proposé par Sen-Yates-Grundy à partir de variables construites au niveau UC qui est finalement retenu en dépit du fait que le plan de sondage indirect de ces unités aboutit à des échantillons de taille variable. Néanmoins, on vérifie que cette variabilité est très faible empiriquement, ce qui permet de rester très proche de l'hypothèse de taille fixe.

En outre, comme elles ne sont pas connues a priori, les probabilités d'inclusion simples et doubles des UC sont estimées par simulation. On semble observer une meilleure robustesse dans l'estimation de ces probabilités par rapport à celles estimées pour les UP. Ceci s'explique en partie par le fait que le taux de sondage des UC est relativement important : un échantillon final de l'EEC contient près du tiers des UC de la population de départ alors que le taux de sondage initial des UP est de l'ordre de  $1/10$ .

### 3.3 Estimation de la variance liée au tirage des secteurs

Après avoir choisi l'estimateur de variance de Sen-Yates-Grundy pour rendre compte de la variance d'échantillonnage des UC, on dispose désormais d'une forme quadratique pour cette première phase comme attendu pour l'utilisation de la formule de Rao donnée par l'équation (1).

L'utilisation de cette expression en tant que formule de variance totale qui articule au mieux les variances de chacune des deux premières phases de tirage repose théoriquement sur l'hypothèse, non vérifiée en pratique, de l'indépendance de tirage des unités secondaires (les secteurs) au sein des unités primaires de première phase (les UC).

On se place malgré tout dans le cadre de cette hypothèse d'indépendance, étant donné que le plan de sondage permet de réaliser un équilibrage, au moins partiel, au niveau des strates régionales. De nouvelles analyses empiriques permettent alors de vérifier que cette hypothèse n'est pas trop forte, en calculant la variance totale à partir de l'expression (1).

En l'absence d'estimateurs de variance conditionnelle, une première validation est faite en utilisant dans un premier temps les variances conditionnelles mesurées empiriquement. On note alors que cette formule de variance totale donne des estimations faiblement biaisées et que l'essentiel de ce biais semble provenir de l'estimateur de variance du tirage des UC.

On note que l'adéquation empirique de la formule de Rao pour une variable d'intérêt donnée, dépend fortement et inversement de l'équilibrage de cette variable que l'on mesure à partir du CV de l'estimateur du total. La figure 1 illustre cette relation inverse entre équilibrage et qualité d'estimation de l'écart-type qui semble correspondre à l'inverse d'une exponentielle.

Pour les variables les plus équilibrées, on s'écarte donc davantage des conditions d'utilisation de la formule de Rao, à savoir l'hypothèse d'indépendance, ce qui conduit à des biais relatifs plus élevés.

En effet, toutes choses égales par ailleurs, les variables auxiliaires servant à l'échantillonnage sont d'autant plus équilibrées qu'elles ont été mobilisées pour arbitrer au niveau régional sur les unités à inclure, après mise en commun de l'ensemble des secteurs des UC appartenant à une strate régionale donnée. La formule de Rao conduit à un biais relativement fort pour estimer la variance relative aux variables que l'on assimile à des variables d'équilibrage et qui sont représentées ici par des croix rouges. En comparaison, la variance relative à des variables d'enquête, représentées par des croix bleues, serait estimée avec un biais relatif beaucoup plus limité.

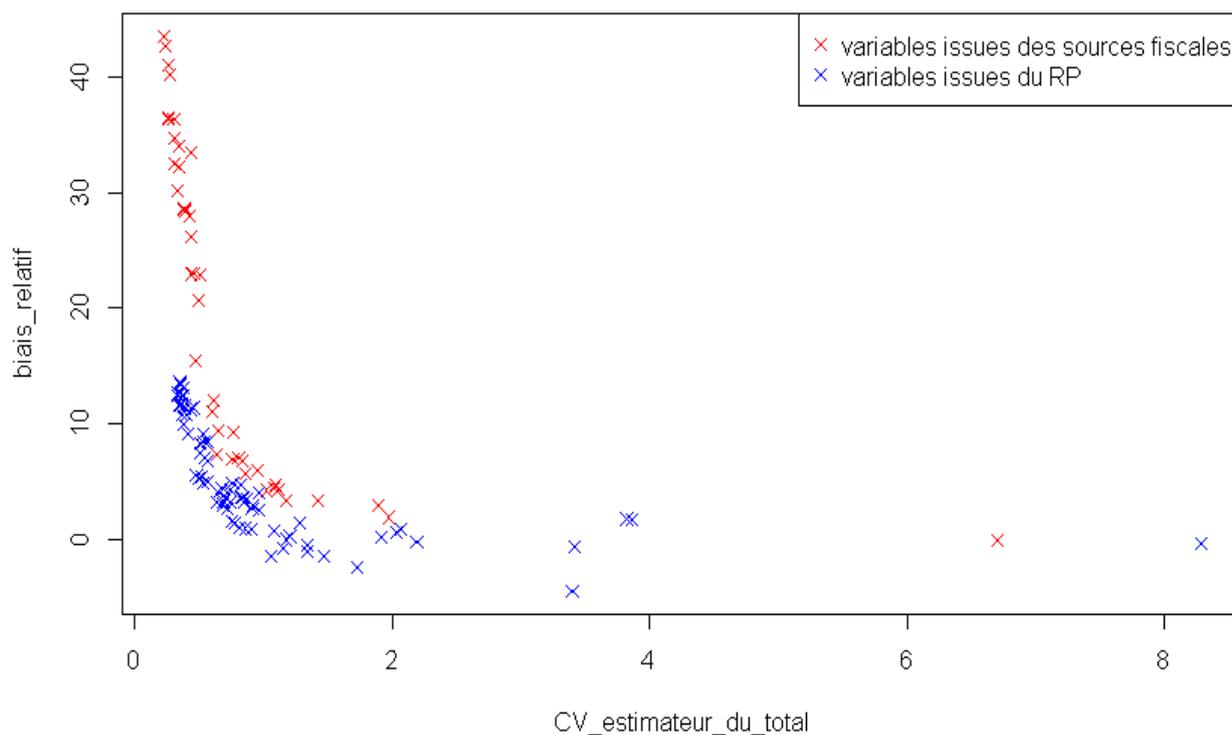


FIGURE 1 – Biais relatif moyen des estimations d'écart-type à partir de la formule de Rao en fonction du CV (%) de l'estimateur du total

La détermination des estimateurs de variance conditionnelle pour le tirage des secteurs constitue le dernier maillon de la chaîne pour l'application de la formule de variance totale de Rao.

Compte tenu du faible nombre de secteurs tirés par UC (entre 4 et 8 secteurs dans la très grande majorité des cas), le nombre de variables d'équilibrage (une quinzaine) est loin d'être négligeable. Ceci restreint fortement la possibilité d'utiliser l'estimateur de Deville-Tillé.

Dans le cas général, c'est donc l'estimateur proposé par Deville qui est utilisé. Cet estimateur étant adapté aux plans de sondages de taille fixe et à probabilités inégales. Il ne tient en revanche pas compte de l'équilibrage, objectif partiellement atteint au sein des UC ; cela conduit donc à des niveaux de biais assez dégradés UC par UC. À partir d'un certain seuil de secteurs tirés au sein d'une UC, les comparaisons empiriques montrent que l'on peut utiliser l'estimateur de Deville-Tillé en ne considérant que quelques variables d'équilibrages (les plus équilibrées) parmi la quinzaine de départ.

Finalement, à l'échelle nationale ou régionale, l'utilisation d'estimateurs de variances conditionnelles en partie biaisés conduit à une amplification limitée du biais dans le calcul de la variance totale de tirage des secteurs.

### 3.4 Conclusion et mise en oeuvre

Les derniers éléments empiriques évoqués à la fin de la section précédente permettent ainsi de valider le recours à la formule de Rao pour estimer la variance due au tirage des secteurs. L'application de cette formule de décomposition de variance se fait par le biais de l'estimateur de Sen-Yates-Grundy pour la variance de premier degré et l'utilisation d'estimateurs de Deville ou Deville-Tillé pour la variance dite de second degré.

## Bibliographie

- [1] Breidt, F.J. and Chauvet (2011), Improved variance estimation for balanced samples drawn via the cube method, *Journal of statistical planning and inference*, No. 1414, pp. 479-487.
- [2] Caron, N., Deville, J-C. et Sautory, O. (1998), *Estimation de précision de données issues d'enquêtes : document méthodologique sur le logiciel POULPE*, Document de travail de l'Insee, Série Méthodologie statistique, No 9806.
- [3] Chauvet, G. (2009), Stratified balanced sampling, *Survey Methodology*, Vol. 35, No. 1, pp. 115-119.
- [4] Costa, L., Merly-Alpa, T. et Chevalier, M. (2018), Le renouvellement de l'échantillon Emploi : améliorations et évolutions, *Actes des Journées de Méthodologie Statistique de 2018*, Insee.
- [5] Deville, J-C. et Lavallée, P. (2006), Sondage indirect : les fondements de la méthode généralisée du partage des poids, *Technique d'enquêtes*, Vol. 32, No 2, p. 185.
- [6] Deville, J-C. and Tillé, Y. (2004), Efficient Balanced Sampling : The Cube Method, *Biometrika*, Vol 91, No 4, pp. 893-912.
- [7] Deville, J-C. and Tillé, Y. (2005), Variance approximation under balanced sampling, *Journal of Statistical Planning and Inference*, Vol 128, pp. 569-591.
- [8] Grafström, A. and Tillé, Y. (2013), Doubly balanced spatial sampling with spreading and restitution of auxiliary totals, *Environmetrics 14*, pp. 120-131.
- [9] Paliot, N., Chevalier, M. et Deroyon, T. (2018), Coordination spatiale d'échantillons : applications à l'EEC et l'Echantillon-Maître, *Actes des Journées de Méthodologie Statistique de 2018*, Insee.
- [10] Rao, J.N.K. (1975), Unbiased variance estimation for multi-stage designs, *Sankhya, C*, 37, pp. 133-139.
- [11] Yates, F. and Grundy, P.M. (1953), Selection without replacement from within strata with probability proportional to size, *Journal of the Royal Statistical Society, Ser. B*, 15, pp. 253-261.