

La coordination d'échantillons des enquêtes auprès des entreprises

Emmanuel Gros & Thomas Merly-Alpa

Résumé — L'objectif de cette note méthodologique est de décrire de façon rapide les nouvelles méthodes mises en œuvre à l'INSEE pour la coordination, positive et négative, entre les échantillons tirés dans le cadre des enquêtes auprès des entreprises.

Cette note reprend très largement l'article d'E. Gros dans la lettre n°73 du SSE de juillet 2016 (document interne INSEE).

I. LA COORDINATION D'ÉCHANTILLONS – PRINCIPE ET PROCÉDURE UTILISÉE À L'INSEE

A. Principes et objectifs

L'objectif de la coordination d'échantillons consiste à prendre en compte, lors du tirage d'un échantillon donné, ceux des enquêtes précédentes, en général dans une optique de réduction de la charge statistique imposée aux entreprises, tout en conservant le caractère sans biais des échantillons.

Deux types de coordination sont distingués. La **coordination négative** consiste à favoriser, lors du tirage d'un échantillon, la sélection des entreprises qui ne l'ont pas été pour des enquêtes récentes. Elle s'inscrit dans une démarche de réduction de la charge statistique imposée aux petites entreprises, les grandes entreprises (à partir d'un certain seuil) étant systématiquement questionnées dans la plupart des enquêtes. À l'inverse, la **coordination positive** vise à maximiser le recouvrement entre les échantillons coordonnés, soit dans une optique de panélisation, soit une fois encore dans l'objectif d'une réduction de la charge statistique globale, par le biais cette fois-ci d'une réduction de la taille des questionnaires d'enquêtes aux thématiques proches (les réponses à une enquête pouvant être utilisées pour une autre sans être collectées via celle-ci).

B. La procédure utilisée à l'INSEE depuis fin 2013

La méthode de coordination en vigueur à l'INSEE [1]¹ permet de tirer des échantillons selon un **plan de sondage aléatoire simple stratifié**² coordonné, et repose sur deux concepts fondamentaux : l'attribution d'un numéro aléatoire permanent à chaque unité de la sphère entreprise, et l'utilisation de fonctions de

coordination. Plus précisément, la procédure de tirage coordonné fonctionne de la façon suivante :

- chaque unité se voit attribuer³ un numéro aléatoire⁴ permanent et invariant qui lui restera attaché tout au long de son existence ;
- lors du tirage d'un échantillon, la sélection des unités s'effectue en retenant dans chaque strate les unités présentant les plus petits numéros aléatoires permanents transformés via une fonction de coordination, déterminée pour assurer la sélection d'un échantillon qui satisfasse à l'objectif de coordination recherché ;
- le choix judicieux de ladite fonction de coordination s'effectue en prenant en compte la fonction de charge cumulée liée aux enquêtes passées, laquelle se calcule à partir des fonctions de coordination utilisées lors des tirages pour ces enquêtes.

Cette procédure de coordination présente de nombreux avantages dans la mesure où elle permet :

- de **coordonner un nombre quelconque d'enquêtes** entre elles. Il est même possible de coordonner l'échantillon positivement avec certaines enquêtes du passé et négativement avec d'autres ;
- de gérer une coordination différenciée selon les enquêtes, en pondérant les tirages dans le processus de coordination en fonction de la durée du questionnaire ou de l'appartenance à un panel par exemple ;
- d'assurer une **coordination entre échantillons de niveaux différents** : groupes, entreprises profilées (EP), unités légales (UL) et/ou établissements. Cette coordination multi-niveaux présuppose une définition de liens univoques et pérennes entre unités de niveaux différents : chaque groupe a une unique EP principale, chaque EP a une unique UL principale, et chaque UL a un seul établissement principal. Les unités ainsi liées se voient affecter le même numéro aléatoire permanent, ce qui permet ensuite d'assurer la coordination multi-niveaux. Cette coordination est donc partielle : dans le cas où une UL a plusieurs établissements par exemple, seul l'établissement principal se voit

1. Les chiffres entre crochets renvoient à la bibliographie à la fin de la note.

2. Il s'agit du plan de sondage standard dans les enquêtes entreprises à l'INSEE

3. Lors de l'initialisation du système pour les unités existant à cette date, ou lors de leur création pour les autres unités.

4. Suivant une loi uniforme sur [0 ; 1]

attribuer la charge de son UL.

II. PERFORMANCES DE LA PROCÉDURE DE COORDINATION

Des études par simulation ont permis de quantifier l'efficacité de cette procédure de coordination en termes de répartition de la charge de réponse entre les différentes unités de la population. La méthode consiste à enchaîner le tirage de vingt échantillons d'UL, chaque échantillon étant alors coordonné négativement avec les échantillons précédents [2]. Le tableau 1 compare la distribution de la variable « charge d'enquête » (définie comme le nombre d'échantillons auxquels une unité appartient) résultant de ces tirages coordonnés avec celle obtenue dans le cadre de tirages indépendants.

Charge d'enquête	Tirages indépendants	Tirages coordonnés
0	3 981 423	3 952 718
1	391 840	445 402
2	30 494	9 084
3	3 670	606
4	374	9
5	18	0

Fig. 1. Charge d'enquête pour 20 tirages indépendants ou coordonnés.

Comme attendu, on observe un resserrement de la distribution autour de 1, i.e. un étalement de la charge d'enquête : le nombre d'unités interrogées plus d'une fois diminue dans des proportions importantes, de même que le nombre d'unités non échantillonnées, au profit d'une augmentation très nette du nombre d'unités sélectionnées dans une seule enquête. Par ailleurs, **cette méthode de coordination s'avère plus performante [3] que l'ancienne méthode utilisée à l'INSEE [4]**, laquelle ne permettait de coordonner les enquêtes que deux à deux.

III. CONTRAINTES INHÉRENTES À LA PROCÉDURE DE COORDINATION

La mise en place de cette procédure de coordination globale impose des limites quant à la coordination des échantillons la première année (ou les X premières années pour les panels rotatifs renouvelés en X fois) où ils rentrent dans le système. En effet, **la coordination ne peut se faire qu'avec des échantillons tirés avec la nouvelle méthode** car il est nécessaire, d'une part d'utiliser le numéro aléatoire permanent pour réaliser le tirage, et d'autre part de disposer des fonctions de coordination ayant servi aux tirages passés pour procéder à la coordination.

Concrètement, cela implique qu'il est impossible de coordonner le tirage de l'échantillon d'une

enquête donnée avec des échantillons d'enquêtes tirés avant l'entrée en vigueur de cette procédure de coordination fin 2013. En particulier, dans le cas de panels rotatifs, la coordination avec les panels passés est impossible lors des renouvellements effectués les premières années. Ainsi, pour tout panel rotatif avec renouvellement par moitié/tiers/quart/etc., la coordination avec les panels passés ne pourra être « complète » que lorsque l'on aura renouvelé tout l'échantillon rotatif par le biais de la nouvelle procédure de coordination.

Par ailleurs, cette procédure de coordination est **incompatible avec la sélection d'échantillons stratifiés avec tirage systématique sur données triées au sein des strates**, qui était auparavant fréquemment utilisée à l'INSEE en vue d'obtenir, au sein de chaque strate, une répartition des unités de l'échantillon proche de celle observée dans la base de sondage pour le critère de tri. Cependant, le tirage systématique s'apparentant à une stratification implicite avec allocations proportionnelles, il est possible de prendre en compte le critère jadis « contrôlé » par le tirage systématique dans la méthode de tirage coordonné en rendant explicite cette stratification, via une méthode de « surstratification » [2].

IV. EXEMPLES D'APPLICATION

A. Enquêtes classiques

Dans la plupart des enquêtes de l'INSEE et du SSP, une coordination négative est assurée avec les autres enquêtes collectées la même année, ceci afin de limiter la charge d'enquête. Cette opération s'inscrit dans le cadre de la simplification de l'action publique [5] qui vise à permettre aux petites entreprises de ne répondre qu'à une seule enquête obligatoire non européenne par an. Les enquêtes sont également souvent coordonnées avec des enquêtes antérieures portant sur des thèmes ou des champs proches, afin de ne pas créer d'effet de lassitude auprès des entreprises.

B. Enquêtes conjoncturelles

Dans le cadre des enquêtes mensuelles de conjoncture, le renouvellement d'échantillons entre les années est toujours une opération complexe, car il faut garantir la stabilité des indicateurs. La méthode de coordination positive permet de garantir un recouvrement optimal des différents échantillons annuels, tout en conservant un caractère sans biais et en permettant une coordination négative avec d'autres enquêtes, afin d'éviter une charge importante de collecte, les enquêtes mensuelles étant souvent assez lourdes pour les entreprises.

Dans ce cas, on ne contrôle cependant pas totalement le nombre d'unités conservées d'un échantillon sur l'autre, ni quelles unités sont réinterrogées. La date de sortie du panel peut

également varier entre les unités.

À l'INSEE, des enquêtes telles que l'EMAGSA (Enquête Mensuelle sur l'Activité des Grandes Surfaces Alimentaires) utilisent cette méthode, dans le but également de créer une fidélisation des entreprises et ainsi faciliter le contact des unités enquêtées.

C. Panels rotatifs

La méthode de coordination exposée ici peut être utilisée dans le cadre de panels rotatifs. En effet, un panel rotatif est un dispositif dans lequel une partie (par exemple la moitié, ou un cinquième) est renouvelée chaque année, les autres parties restant inchangées. Cela assure une stabilité de l'échantillon que l'on peut mesurer. La coordination négative avec l'échantillon à la date d'entrée dans le dispositif permet alors ici de faire en sorte que les unités « sortantes » ne soient pas réintégréées dans le panel pour un nouveau cycle ; cela assure de limiter la charge de collecte sur les unités en longitudinal, mais aussi de couvrir plus efficacement toute la population d'intérêt.

Les enquêtes TIC (Technologies de l'Information et de la Communication) ainsi que l'ESA (Enquête Sectorielle Annuelle), jusqu'en 2016, utilisent cette méthode afin de disposer à terme d'informations structurelles sur la totalité des entreprises françaises, tout en assurant une cohérence des résultats d'une année sur l'autre.

D. Cas plus atypiques

Enfin, dans certains cas, la méthode de coordination est mise en œuvre afin de respecter des contraintes qui peuvent paraître contradictoires. Dans le cadre d'une extension de l'enquête EACEI (Enquête Annuelle sur les Consommations d'Énergie dans l'Industrie) afin de diffuser des résultats précis à un niveau régional, la question de la coordination avec l'enquête Antipol (investissements Antipollution), sur le même champ et les mêmes thématiques, s'est posée. En effet, les tailles d'échantillons étaient si importantes par rapport au champ, qu'une coordination négative entre EACEI, son extension et Antipol conduisait à réinterroger les mêmes unités. Une étude spécifique a donc dû être menée pour déterminer dans quelles strates le tirage était à coordonner positivement et dans quelles autres il était à disjointer.

REFERENCES

- [1] Fabien Guggemos et Olivier Sautory, « Sampling Coordination of Business Surveys Conducted by INSEE », Proceedings of the Fourth International Conference of Establishment Surveys, Montréal, Canada, 11-14 juin 2012 (<https://ww2.amstat.org/meetings/ices/2012/papers/302211.pdf>).
- [2] Emmanuel Gros, « Études et simulations relatives à la nouvelle méthode de coordination des échantillons d'enquêtes entreprises et établissements développée à l'Insee », Actes des 12e Journées de méthodologie statistique, 31 mars – 02 avril 2015, INSEE.

- [3] Anaïs Levieil-Guillon, « Mise en œuvre et résultats de la nouvelle procédure de coordination des échantillons des enquêtes auprès des entreprises et/ou des établissements », Actes des 12e Journées de méthodologie statistique, 31 mars – 02 avril 2015, INSEE.
- [4] Arnaud Fizzala et Olivier Sautory, « La coordination négative d'échantillons », Lettre du SSE n° 66, 1er semestre 2011, document interne INSEE.
- [5] Circulaire du 16 octobre 2015 relative à la mise en œuvre des mesures de simplification administrative en matière d'enquêtes statistiques pour les petites entreprises, Bulletin officiel de l'administration centrale n° 2015-11 du 5 novembre 2015, <http://www.economie.gouv.fr/bulletin-officiel/boac2015-11>.



*Département des méthodes statistiques
Version n°1, diffusée le 10 octobre 2017*