

Études et simulations relatives à la nouvelle méthode de coordination des échantillons d'enquêtes entreprises et établissements développée à l'Insee

Emmanuel Gros¹

¹Insee, DMCSI, 18 bd A. Pinard, 75675 Paris Cedex 14, emmanuel.gros@insee.fr

Thème : échantillonnage

Le système statistique public réalise chaque année un nombre important d'enquêtes auprès des entreprises et des établissements. L'objectif de la coordination négative d'échantillons est de favoriser, lors du tirage d'un échantillon, la sélection d'entreprises n'ayant pas déjà été sélectionnées lors d'enquêtes récentes, tout en conservant le caractère sans biais des échantillons. Elle s'inscrit donc dans une démarche de réduction de la charge statistique imposée aux petites entreprises – les grandes entreprises, à partir d'un certain seuil, étant systématiquement enquêtées dans la plupart des enquêtes. La coordination positive vise quant à elle 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 par le biais cette fois-ci d'une réduction de la taille des questionnaires.

Lors des onzièmes Journées de Méthodologie Statistique en 2012, Fabien Guggemos et Olivier Sautory avaient présenté dans [1] une nouvelle méthode de coordination des échantillons des enquêtes auprès des entreprises, alors en cours d'expérimentation à l'Insee. Cette méthode – qui fait partie de la famille des procédures de coordination s'appuyant sur l'utilisation de numéros aléatoires permanents – repose sur la notion de *fonction de coordination*. Ces fonctions, définies pour chaque unité et chaque nouveau tirage en fonction des charges de réponse aux enquêtes passées, transforment les nombres aléatoires permanents des unités de façon à assurer la sélection d'un échantillon qui satisfasse à l'objectif de coordination, en général négative, recherché tout en respectant le plan de sondage voulu. Les résultats des premières évaluations empiriques de la méthode menées sur des données simulées et présentés dans [1] se sont révélés très satisfaisants : la nouvelle méthode de coordination s'est ainsi révélée à la fois très efficace – en conduisant à des gains considérables¹ en termes de répartition de la charge de réponse sur les différentes unités de la population – et remarquablement robuste vis à vis des paramètres des différents plans de sondage – taux de sondage, stratification, taux de recouvrement entre le champ des différentes enquêtes, charges associées aux différentes enquêtes, etc.

Le présent article reprend rapidement les fondements théoriques de cette méthode de coordination, en se limitant au cas du tirage aléatoire simple stratifié, qui est l'échantillonnage le plus souvent utilisé à l'Insee pour les enquêtes-entreprises. Il aborde ensuite deux problèmes posés par la méthode de coordination – biais de rétroaction² et incompatibilité de la procédure avec l'algorithme de tirage systématique – et expose les options retenues sur ces deux questions. Il s'intéresse en outre à la question de la coordination entre échantillons de niveaux différents et détaille comment la méthode proposée permet de coordonner des échantillons d'enquêtes portant sur différents types d'unités « emboîtés » – établissements, unités légales, entreprises, groupes – dans le cadre d'un système de coordination « intégré ». Enfin il présente les résultats de nombreuses simulations effectuées sur données réelles en grandeur nature pour évaluer la performance, en termes de répartition de la charge de réponse sur les différentes unités de la population, de la procédure de coordination dans différentes configurations.

Référence :

[1] Fabien Guggemos et Olivier Sautory, « La coordination d'échantillons d'enquêtes auprès des entreprises mise en place à l'Insee », actes des XIèmes Journées de Méthodologie Statistique, 2012.

¹ Par rapport à des tirages indépendants.

² Biais qui apparaît lorsque l'on coordonne une enquête B avec une enquête A dont les résultats ont servi à mettre à jour la base de sondage de l'enquête B.