

Analyse et traitement du « hors champ » Cas de l'enquête Acemo-TPE

Ludovic VINCENT (Dares)

Les sources disponibles pour constituer la base de sondage d'une enquête ne permettent pas toujours de circonscrire exactement la population d'intérêt. Certaines unités statistiques peuvent y être indûment présentes et donc sélectionnables dans l'échantillon à construire, constituant alors une population hors du champ pour l'enquête.

Les raisons tiennent à la longueur des processus de mise à jour des bases de sondage au regard de la démographie des entreprises, ce qui engendre une différence entre l'image du tissu productif de la base et la réalité que souhaite mesurer le statisticien à une date donnée.

Ce problème est particulièrement prégnant pour les enquêtes dont le champ est défini par des seuils, comme l'enquête Acemo Très Petites Entreprises (TPE), laquelle est administrée auprès des entreprises de 1 à 9 salariés. Les mouvements de main d'œuvre au sein des entreprises sont tels que les passages de 9 à 10 salariés ou plus, ou ceux de 1 à 0 salarié sont en effet très fréquents.

Pour ce type d'enquête, il est donc particulièrement nécessaire de traiter attentivement le hors champ en raison de son importance (13 % de l'échantillon dans le cas de l'enquête TPE)). Si des unités statistiques se signalent d'elles-mêmes hors champ au cours de la collecte (10 % des répondants, soit 6 % de l'ensemble de l'échantillon), un nombre important le sont parmi celles ne retournant pas le questionnaire (7-8 % de l'échantillon). Elles le demeurent en général au moment où l'exploitation de l'enquête est effectuée et ces unités sont considérées comme non répondantes et indûment dans le champ, biaisant les résultats.

L'étude recourt à des données supplémentaires (sources auxiliaires et/ou base plus récente) pour détecter ces unités pour l'enquête Acemo TPE, permettant de quantifier le phénomène, et les caractériser.

Ces sources additionnelles mises à disposition postérieurement à la date de constitution de la base de sondage permettent de repérer la majeure partie de cet ensemble d'entreprises (soit 7 % de l'ensemble de l'échantillon). Ces sources contiennent une information plus adéquate que celle utilisée au moment de la constitution de la base de sondage.

Pour traiter les unités statistiques pour lesquelles le statut hors champ est indéfini (1 % de l'échantillon), l'étude compare différentes méthodes statistiques : repondération ou modèles statistiques prédictifs.

Enfin, l'impact sur les estimateurs produits du repérage ex-post d'unités hors champ (7-8 % de l'échantillon) est mesuré pour quelques estimateurs habituellement publiés.