
CORRECTION DE LA NON RÉPONSE ENDOGÈNE : UNE APPLICATION SUR L'ENQUÊTE POST-FORMATION

Alejandra Arbelaez Ayala (), Anne Bucher (*), Melchior Fosse (**) et Pauline Givord (*)*

() Dares, Pôle Evaluation du Plan d'investissement dans les compétences*

*(**) Insee, Division Salaires et revenus d'activités (SRA)*

Pauline.givord@travail.gouv.fr

Mots-clés: emploi, formation, non réponse endogène

Domaine concerné : Multimode, Redressement, pondération et repondération, calage sur marges

Résumé

L'enquête Post-Formation » a été lancée en 2019 par la Dares pour contribuer à l'évaluation du Plan d'Investissement dans les Compétences. Elle vise à recueillir de l'information auprès des stagiaires de la formation professionnelle sur le déroulement de leur formation, leurs motivations et leurs attentes, ainsi que leurs appréciations sur le contenu des formations reçues.

Il s'agit d'une enquête menée chaque trimestre auprès de personnes ayant été en recherche d'emploi et ayant, à ce titre, suivi une formation professionnelle. L'enquête cible les personnes entre 6 et 9 mois après leur sortie de formation, y compris celles qui l'ont interrompue avant la fin. L'échantillon est constitué à partir des fichiers administratifs des stagiaires de la formation professionnelle (BREST), avec un objectif de 25 000 à 30 000 répondants par trimestre, soit 2 000 répondants par région de France métropolitaine hors Corse et en moyenne 1 000 pour la Corse et les départements d'Outre-Mer.

L'enquête est auto-administrée, avec un recueil sur internet ou sur papier. Les taux de réponse sont relativement faibles (autour de 35% pour chaque vague), ce qui peut poser des questions sur la qualité des estimateurs obtenus, en présence de non réponse endogène. C'est notamment le cas du comportement d'abandon en cours de formation, un indicateur important pour le suivi.

L'estimation du taux d'abandon peut être biaisée, si les stagiaires qui ont abandonné ont également une probabilité plus faible de répondre à l'enquête, même conditionnellement aux variables disponibles dans la base d'échantillonnage et utilisée pour le redressement.

Néanmoins, sur certaines vagues de l'enquête, un sur-échantillonnage a été mis en place, avec l'utilisation de relances téléphoniques pour les personnes ne répondant pas au questionnaire auto-administré. Comme montré par Castel et Sillard (2021), il est alors possible de redresser

des biais de non réponse endogène en appliquant un modèle de sélection endogène d'Heckman (Tobit II). On modélise simultanément la probabilité d'avoir abandonné la formation et le comportement de réponse à l'enquête, à partir d'un modèle de sélection où le fait d'appartenir à l'échantillon avec relance est utilisé comme instrument.

L'application de cette méthode à l'enquête Post Formation confirme que les personnes qui ont abandonné sont aussi plus souvent non répondant à l'enquête, ce qui entraîne une sous-estimation du taux d'abandon en l'absence de correction. La correction du biais de non réponse par cette méthode se traduit cependant par une perte de précision des estimateurs, perte qui se réduit avec la taille de l'échantillon de relance. Un arbitrage est donc à opérer sur le plan opérationnel pour maintenir un niveau de précision suffisante, avec un niveau de coût acceptable.

Bibliographie

[1] L. Castell and P. Sillard (2021): Le traitement du biais de sélection endogène dans les enquêtes auprès des ménages par modèle de Heckman, document de travail de la DMCSI N° M2021/02, INSEE

[2] J.J. Heckman. Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, pages 153–161, 1979.