
LE TRAITEMENT DU BIAIS DE SÉLECTION ENDOGÈNE DANS LES ENQUÊTES AUPRÈS DES MÉNAGES PAR MODÈLE DE HECKMAN

Laura Castell, Patrick Sillard ()*

() Insee, Direction de la méthodologie et de la coordination statistique et internationale
laura.castell@insee.fr ; patrick.sillard@insee.fr*

Mots-clés : sélection endogène, multimode, instrument, modèle d'Heckman

Domaine concerné : redressement ; multimode

Résumé

Dans la plupart des enquêtes auprès des ménages, les méthodes de correction de la non-réponse mises en œuvre corrigent la sélection sur observables. Elles font en effet l'hypothèse que le fait de répondre à l'enquête – lié pour partie à l'intérêt pour la thématique de l'enquête – est un mécanisme ignorable (missing at random, ou MAR), c'est-à-dire que la non-réponse est indépendante des variables d'intérêt, conditionnellement aux caractéristiques (observées dans la base d'enquête) prises en compte dans le modèle de correction.

Dans un certain nombre d'enquêtes auprès des ménages, on peut suspecter un phénomène de sélection endogène, c'est-à-dire que des déterminants des variables d'intérêt – tels que l'intérêt pour la thématique de l'enquête – influent également sur la participation, conditionnellement aux caractéristiques observables. Dans ce cas, le mécanisme de non-réponse est non-ignorable et les méthodes usuelles de correction de la non-réponse conduisent à des estimations biaisées. Ce biais est d'autant plus important que le taux de réponse est faible et que la variable omise est corrélée aux variables d'intérêt.

Dans le cadre du développement de la collecte par Internet dans les enquêtes auprès des ménages, cette problématique de la sélection endogène devient a priori plus prégnante. De fait, le mécanisme de réponse par Internet nécessite une démarche plus pro-active que des modes de collecte intermédiés par un enquêteur. Il est donc plus probable que la non-réponse fasse davantage l'objet d'une auto-sélection, très liée à l'intérêt pour la thématique de l'enquête, et que cet intérêt ne soit pas entièrement corrélé à des caractéristiques observables. Ce biais éventuel risque d'être d'autant plus marqué que les taux de réponse des enquêtes par Internet restent relativement faibles.

Dans cet article, nous étudions les conditions dans lesquelles une correction de la sélection endogène peut être réalisée et en particulier quelle configuration d'enquête autorise la mise en œuvre d'une correction.

Les méthodes examinées sont basées sur le modèle d'Heckman et nécessitent de disposer d'instruments issus d'un protocole de collecte adapté en amont.

Usuellement, en économétrie, la correction de la sélection endogène est réalisée par imputation de la variable d'intérêt pour les non-répondants. De manière similaire, on peut réaliser cette correction

par repondération en utilisant la probabilité d'inclusion conditionnelle $E(r_i|z_i, w_i, y_i)$ du modèle. Cette méthode inédite correspond davantage aux usages de correction de la non-réponse dans la production d'enquêtes auprès des ménages dans la statistique publique. C'est celle qui est développée dans l'article.

La mise en œuvre d'une telle méthode suppose l'existence d'un instrument expliquant la participation et non la variable d'intérêt, et vérifiant le principe de monotonie.

De tels instruments peuvent être créés en adaptant les protocoles d'enquêtes en amont de la collecte. Une façon de procéder consiste à décomposer l'échantillon d'enquête en plusieurs sous-échantillons tirés aléatoirement et administrés selon des protocoles différents impliquant des taux de participation distincts. L'utilisation de différents modes de collecte fait partie des pistes d'adaptation de protocoles pour se retrouver dans une telle situation instrumentale.

Au-delà de la correction de la sélection endogène, les effets de mode peuvent survenir en lien avec l'existence d'erreur de mesure liées au mode de réponse lui-même. Dans ce cas, erreur de mesure et sélection endogène se confondent et séparer les deux peut se révéler compliqué.

L'article discute différents cas de figure et pistes envisageables pour traiter des effets de mode (sélection endogène et erreur de mesure) dans des enquêtes mobilisant des protocoles multimodes.

Bibliographie

- [1] Castell L., Sillard P., « Le traitement du biais de sélection endogène dans les enquêtes auprès des ménages par modèle de Heckman », *Documents de travail*, n°M2021/02, Insee, 2021.
- [2] Heckman J.J., « Sample selection bias as a specification error », *Econometrica*, vol 45, n°1, p. 153-161, 1979.