

Biais de sélectivité portant sur des variables quantitatives sur données longitudinales.

Stéfan LOLLIVIER

INSEE, Directeur des Statistiques Démographiques et Sociales

Depuis Heckman (1978), on a pris conscience de l'importance des biais que peuvent introduire des mécanismes de sélection dans l'estimation de comportements décrits au travers de variables quantitatives (salaires féminins,...). La contribution d'Eric Gautier aux JMS 2005 rappelle comment des procédures d'estimations relativement économes en matière de restrictions paramétriques sur les lois des termes d'erreur peuvent être mises en œuvre au moyen de procédures disponibles directement dans les logiciels usuels.

On se propose ici de généraliser la formalisation d'Heckman aux données longitudinales. Le formalisme développé par Wooldridge (1995) part du principe que la variable observée comme la variable de sélection peuvent être décrites au moyen de modèles à effets individuels stochastiques. Moyennant un jeu de quatre hypothèses, Wooldridge aboutit à une réécriture des équations. Deux d'entre elles portent sur la spécification de l'équation de sélection, les deux autres sur la variable d'intérêt. Comme chez Heckman, les équations qui résultent de ces hypothèses peuvent être estimées au moyen de modèles probit simples en ce qui concerne la sélection ou des moindres carrés ordinaires pour l'équation d'intérêt. L'estimateur ainsi obtenu est asymptotiquement convergent et normal, mais sa matrice de variance-covariance asymptotique est complexe à calculer. On peut cependant l'obtenir au moyen de méthodes de bootstrap ; la technique du bootstrap par paires (Efron et Tibshirani, 1986) est bien adaptée à la nature du problème, et très simple à mettre en œuvre. De ce fait, la méthode de Wooldridge est accessible à tous. D'autres méthodes sont disponibles dans la littérature. Plus économes en hypothèses, elles sont cependant plus complexes à mettre en œuvre, du fait de recours quasi-inévitables à des procédures de lissage, par exemple au moyen de noyaux.

Références

Efron B., et R. Tibshirani (1986): « Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy », *Statistical Science*, 1, n°1, 54-77.

Heckman J. (1978): « Dummy endogenous variables in a simultaneous equation system », *Econometrica*, 46, 931-959.

Wooldridge J. M. (1995) : « Selection corrections for panel data models under conditional mean independence assumptions », *Journal of Econometrics*, 68, 115-132.