

NON RÉPONSE À L'ENQUÊTE EMPLOI ET MODÈLES PROBIT SPATIAUX

Vincent Loonis ¹

¹ *Insee, 18 bd Adolphe Pinard, 75675 PARIS CEDEX 14,*
`vincent.loonis@insee.fr`

Préfigurant certaines opportunités associées au futur répertoire statistique des logements de l'Insee, l'échantillon de l'Enquête Emploi en Continu (EEC) est depuis le dernier trimestre de 2010 entièrement sélectionné dans des fichiers administratifs, en l'occurrence ceux de la taxe d'habitation.

La pérennité des identifiants autorise la description de l'ensemble des logements échantillonnés (répondants et surtout non-répondants) par des caractéristiques observées au moment de l'enquête. Les descripteurs concernent le logement lui-même, le ou (les) ménage(s) y résidant (profil sociodémographique et structure des revenus) mais aussi sa localisation géographique fine (X,Y).

Il est dès lors possible d'envisager un enrichissement de l'étude du mécanisme de non-réponse à l'EEC. On peut en particulier documenter simultanément la présence de 3 effets :

- l'impact d'un plan de sondage aréolaire sur la propension de ménages à répondre,
- la présence d'un effet enquêteur,
- la non-ignorabilité du mécanisme de non réponse.

On mobilise pour cela des modèles probit spatiaux. Ces modèles intègrent la géographie au travers de matrices de contiguïté. Dans notre cas, ces matrices peuvent être simples : deux logements sont contigus s'ils appartiennent à la même aire, ou plus raffinées. En s'appuyant sur les coordonnées X,Y on peut, par exemple, définir la contiguïté par triangulation de Delaunay.

L'estimation des modèles probit spatiaux est généralement complexe et nécessite le recours, soit au maximum de vraisemblance simulé, soit à des méthodes bayésiennes type MCMC. Dans notre cas, la spécificité des matrices de contiguïté permet de mobiliser le pseudo maximum de vraisemblance dont la contrepartie empirique est très simple : elle se limite à estimer un probit simple sur données empilées (Lollivier, 2009). La précision des estimations est obtenue par bootstrap pour tenir compte de la nature autorégressive du modèle et de l'autocorrélation des résidus qu'elle implique.

Bibliographie

Lollivier, S. (2009). Pseudo-maximum de vraisemblance : comment estimer simplement des modèles qualitatifs statiques avec variables endogènes sur données de panel, article présenté aux Journées de Méthodologie Statistique, Paris, France.