

Traitement de la non-réponse non-ignorable par calage généralisé: une simulation à partir de l'enquête Budget des Ménages au Luxembourg

Guillaume OSIER¹

Dans la plupart des cas, malgré les efforts déployés par les gestionnaires d'enquête pour l'anticiper, la non-réponse totale dans les enquêtes ménages ne peut être complètement éradiquée.

C'est pourquoi on a souvent recours à des techniques de repondération afin de corriger le biais de non-réponse, dû à des différences entre le profil moyen des ménages répondants et celui des ménages non-répondants. A ce niveau, une information auxiliaire corrélée avec le fait pour un ménage de répondre ou pas à l'enquête est nécessaire afin de réduire le biais de non-réponse de manière significative. Malheureusement, la disponibilité d'une telle information est souvent problématique.

Les méthodes de repondération pour la non-réponse cherchent souvent à estimer la probabilité de réponse des ménages qui ont participé à l'enquête.

La technique courante est d'utiliser un modèle logistique avec comme variables explicatives des variables dont les valeurs sont connues à la fois sur les ménages répondants et sur les ménages non-répondants. Ce dernier point est une sérieuse contrainte dans la pratique, notamment lorsque la non-réponse dépend du sujet-même de l'enquête (mécanisme non-ignorable). Dans ce cas, les variables auxiliaires les plus corrélées avec la non-réponse sont uniquement connues sur les ménages répondants. On n'est donc pas en mesure d'estimer la probabilité de réponse des ménages.

Le recours aux techniques de calage pour traiter la non-réponse totale permet de s'affranchir de cette contrainte, dans la mesure où les variables que l'on utilise pour réaliser le calage n'ont à être connues que sur les unités répondantes. Cependant, l'information sur la valeur des totaux de calage est requise, ce qui constitue encore une sérieuse contrainte dans la pratique.

L'approche par calage généralisé (Deville, 2002 ; Le Guennec et Sautory, 2002 ; Ardilly, 2006) permet de surmonter cette dernière difficulté en distinguant dans les équations de calage les variables intervenant pour la réduction du biais de non-réponse de celles intervenant pour le calage proprement dit, c'est-à-dire la réduction de la variance d'échantillonnage. Pour les premières, la seule condition est de connaître leurs valeurs sur les ménages répondants, tandis que pour les dernières il faut en plus connaître la valeur des totaux sur lesquels on cale. L'avantage de cette forme de calage est que l'on peut intégrer dans notre modèle de réponse des variables qui ont été collectées pendant l'enquête, et

¹ *Auteur:* M. Guillaume Osier

Affiliation : Institut National de la Statistique et des Etudes Economiques du Grand-duché de Luxembourg (STATEC)

Adresse :

Centre Administratif Pierre Werner

13, rue Erasme

B.P. 304

L-2013 Luxembourg

G.D. Luxembourg

Tel : (+352) 247 84374

Courriel : Guillaume.Osier@statec.etat.lu

dont on ne connaît que les valeurs sur les unités répondantes. Lorsqu'on est en présence d'un mécanisme non-ignorable, ces variables sont fortement corrélées avec la non-réponse. Ce papier présente les résultats d'une simulation réalisée à partir des données de l'enquête Budget des Ménages (EBM) au Luxembourg.

L'échantillon utilisé pour la simulation comprend près de 8000 ménages qui ont répondu à l'enquête entre 2005 à 2010. On a tiré 2000 répliques selon un mécanisme non-ignorable attribuant une probabilité de tirage plus faible aux ménages qui ont une dépense élevée. Si les pondérations ne tiennent pas compte du caractère non-ignorable du mécanisme de tirage de l'échantillon, un biais important va s'en suivre dans nos estimations.

La technique du calage généralisé implantée dans la nouvelle version de la macro Calmar (Calmar2) a alors été testée comme moyen permettant de corriger le biais de non-réponse.

Références

Ardilly, P. (2006). *Les techniques de sondage*. Technip.

Deville, J.C. (2002). *La correction de la non-réponse par calage généralisé*. Actes des JMS 2002.

Le Guennec, J. et Sautory, O. (2002). *Application du calage généralisé à la correction de la non-réponse : une expérimentation*. Actes des JMS 2002.