
COMMENT UTILISER DES DONNÉES D'ENQUÊTE ? GUIDE ILLUSTRÉ SUR L'ÉCONOMÉTRIE DES DONNÉES ÉCHANTILLONNÉES

Cyril FAVRE-MARTINOZ (*), Marine GUILLERM (**), Ronan LE SAOUT (***)

(*) Insee, DR de La Réunion

(**) Dares

(***) Crest-Ensaï

cyril.favre-martinoz@insee.fr
marine.guillerm@travail.gouv.fr
ronan.le-saout@ensai.fr

Mots-clés : poids de sondage, enquête statistique, sélection

Résumé

Les données d'enquête sont largement utilisées pour mener des études économiques. Les liens entre la micro-économétrie et la théorie des sondages sont pourtant mal connus ou mal compris, ces deux pans de la statistique ayant des objectifs différents : inférence des paramètres d'un modèle vs estimation de statistiques descriptives. Les chargés d'études font souvent abstraction du fait qu'ils utilisent des données d'enquête. Ils ne savent pas toujours comment tenir compte des traitements d'enquête (plan de sondage et pondération, repondération liée à la non-réponse ou au calage). Nous conduisons ici une revue de littérature sans se limiter à la théorie économétrique. Les contraintes tant en termes de logiciels que de contenu des bases de données sont ainsi soulignées à travers des exemples basés sur l'enquête Patrimoine conduite par l'INSEE. La question de la pondération ou non des modèles économétriques a été largement abordée dans la littérature. Il n'est pas nécessaire de pondérer lorsque les variables ayant servi à définir le plan de sondage sont incluses dans le modèle. Dans les autres cas, une approche prudente reste de comparer les estimations des modèles pondérés et non pondérés. Pour le calcul de variance tenant compte du plan de sondage, deux types d'aléa doivent être pris en compte, l'aléa du tirage de l'échantillon et celui du modèle économétrique. Le calcul précis de cette variance nécessite ainsi des informations détaillées sur la méthodologie d'enquête et n'est que rarement effectué. Dans de nombreux cas, cette variance peut néanmoins être approchée. Les méthodes d'imputation utilisées pour corriger la non-réponse ont des effets ambigus sur les modèles économétriques, ce qui reste encore un sujet à explorer. Le peu de prise en compte de la méthodologie d'enquête peut ainsi provenir d'une absence de solution théorique identifiée mais également d'un manque de communication sur les traitements mis en oeuvre lors d'une enquête statistique. La diffusion d'informations détaillées sur la construction des données apparaît donc nécessaire pour permettre une véritable économétrie des données d'enquête.