

La qualité des données sur les revenus : enquête versus fichier administratif

L'exemple de l'enquête SRCV 2005

Laurence DAUPHIN¹

Lorsque le dispositif SRCV a été mis en place en 2004, la source de référence nationale pour la mesure des revenus était, et est toujours, l'enquête Revenus fiscaux (devenue depuis l'enquête Revenus fiscaux et sociaux). La France dispose ainsi aujourd'hui de deux sources de référence sur les revenus : l'une au niveau national, l'enquête Revenus fiscaux et sociaux, l'autre au niveau européen, le dispositif SILC-SRCV. Réconcilier au mieux les deux dispositifs est un enjeu important. Compte tenu des différences de nature entre données administratives et données déclarées dans le domaine des revenus, l'appariement des données SRCV avec les déclarations fiscales est la voie qui paraît la plus réaliste. Ceci doit permettre de pouvoir utiliser les avantages de chacun des deux types de sources, et ainsi d'améliorer la qualité du dispositif en profitant des atouts des sources administratives.

Celles-ci offrent en effet une qualité homogène et très satisfaisante des données sur les revenus pour la grande majorité de la population. Ce n'est pas le cas pour les sources déclaratives. En effet, les erreurs entre euros et francs, voire même entre euros et anciens francs, et entre montants annuels et montants mensuels sont courantes, comme le montrent les travaux d'imputation menés sur les données des premières vagues du dispositif SRCV. Par ailleurs, certains ménages utilisent des documents, d'autres non. Dans le cas où les documents ne sont pas utilisés, les données sont souvent approximatives ou simplement données en tranches. Enfin, une fraction des ménages ne souhaite répondre à aucune question relative aux revenus. Tous ces éléments conduisent à une hétérogénéité dans la collecte qui nuit aux objectifs de qualité globale, et aux objectifs même du panel qui est de suivre d'une année sur l'autre le revenu d'une même personne (Lollivier et Verger, 2005)*.

Cependant, les sources administratives ne permettent pas non plus une mesure parfaite des composantes de revenu. Certaines composantes sont exonérées d'impôts et ne sont pas présentes dans les déclarations fiscales. D'autres composantes sont très mal appréhendées, parce que les montants que l'on peut trouver dans les déclarations fiscales résultent de l'application de barèmes fiscaux, ne correspondant pas à une notion de revenu au sens classique du terme.

Un appariement à visée méthodologique entre les données de la collecte SRCV réalisée en mai-juin 2005 et les déclarations fiscales portant sur les revenus 2004 a été effectué. Cet appariement concerne 5 600 ménages avec un taux de succès de 95 %.

¹ DSDS, Insee - Laurence.dauphin@insee.fr

* Lollivier S. et Verger D., 2005, Trois apports des données longitudinales à l'analyse de la pauvreté, Économie et Statistique n^{os} 383-384-385

Au-delà de la question de la faisabilité, le test a été réalisé afin d'évaluer plus précisément la qualité des données d'enquête et des données administratives obtenues au travers de la déclaration fiscale. Ce test permet de mieux connaître les avantages et inconvénients de ces différentes sources en fonction des variables et populations étudiées. Par ailleurs, il permet également de mettre en lumière les difficultés ainsi que les opportunités qu'offrent les appariements.