

Esane ou les malheurs de l'estimation composite : comment gérer les valeurs négatives d'estimateurs par différence ?

Emmanuel GROS¹

Le nouveau dispositif de production des statistiques structurelles d'entreprises françaises, Esane (Élaboration des Statistiques ANnuelles d'Entreprises), a été mis en place en 2009. Il s'appuie sur une utilisation intensive de sources administratives (déclarations annuelles sur les bénéficiaires adressées par les entreprises à la direction générale des Impôts, déclarations annuelles de données sociales), complétées par des données obtenues par une enquête statistique réalisée sur un échantillon d'entreprises.

Cette utilisation conjointe de données administratives et de données d'enquête intervient à différentes étapes du processus d'exploitation des données – correction de la non-réponse, calage – et en particulier lors de la phase d'estimation via l'utilisation d'estimateurs par différence.

Ce type d'estimateur se révèle particulièrement adapté au contexte du dispositif Esane, puisqu'il permet de mobiliser conjointement les sources administratives et les résultats d'enquête, et présente de nombreux avantages, en particulier en terme de précision. Il ne garantit toutefois pas la positivité des estimations et peut donc conduire à des estimations négatives, alors même que toutes les données individuelles sont positives ou nulles. Ceci se révèle problématique, d'autant plus lorsque cela concerne des variables pour lesquelles des agrégats négatifs n'ont aucun sens économique.

Si ces cas problématiques s'avèrent assez rares tant qu'on raisonne à un niveau suffisamment agrégé, le nombre d'estimations négatives s'accroît sensiblement dès que l'on s'intéresse à des agrégats à un niveau fin de la nomenclature, ou portant sur des variables à occurrences rares. Il a donc été nécessaire d'adapter la procédure d'estimation afin de pallier ce problème.

La gestion de ces estimations négatives problématiques a été rendue d'autant plus difficile par la richesse et la complexité du système Esane : les statistiques produites par ce système sont en effet soumises à de nombreuses contraintes de cohérence, tant « verticales » – cohérence des estimations portant sur différents niveaux de nomenclature hiérarchiquement imbriqués – que « horizontales » – cohérence lors de l'estimation de variables liées entre elles par des relations comptables –, que la méthode d'estimation se doit de respecter.

Cet article reviendra dans un premier temps sur les estimateurs par différence mis en œuvre dans le système Esane, ainsi que sur le problème que posent les estimations négatives, avant de détailler la méthode retenue pour gérer ces estimations problématiques.

¹ Insee (emmanuel.gros@insee.fr)