

# LE DISPOSITIF ESANE, OU COMMENT L'UTILISATION COMBINÉE DE DONNÉES ADMINISTRATIVES ET DE DONNÉES D'ENQUÊTE PERMET D'AMÉLIORER LA QUALITÉ DES DONNÉES INDIVIDUELLES ET DES STATISTIQUES

Emmanuel Gros <sup>1</sup>

<sup>1</sup> *Insee, 18 bd Adolphe Pinard, 75675 PARIS CEDEX 14, emmanuel.gros@insee.fr*

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 (cf. Brion 2009 et Gros 2012 pour plus de détails). Cette utilisation conjointe de données administratives et de données d'enquête intervient à différentes étapes du processus d'exploitation des données et ouvre de nouvelles perspectives, tant en termes de contrôle des données individuelles que lors de la phase d'estimation, via la mise en œuvre de procédures de calage et l'utilisation d'estimateurs composites spécifiques.

En ce qui concerne l'amélioration de la qualité des données individuelles, une procédure de contrôle de cohérence entre sources administratives et données d'enquête a été mise en œuvre pour un petit groupe de variables communes aux différentes sources : chiffre d'affaires, partage du chiffre d'affaires entre ventes de marchandise, production de biens et production de services, effectifs salariés et salaires. À l'issue de cette phase de réconciliation des données individuelles, qui peut conduire au rappel de certaines entreprises, on dispose donc, pour chaque variable concernée, d'une variable réconciliée contenant la valeur finale de référence de cette variable.

Ces variables réconciliées ne sont bien évidemment disponibles que pour les unités de l'échantillon, de même que d'autres variables-clef du dispositif Esane, tel le code APE. Afin de prendre en compte ces informations disponibles uniquement pour les unités de l'enquête tout en mobilisant au mieux l'intégralité des informations administratives, une procédure d'estimation composite spécifique, reposant essentiellement sur des estimateurs par différence, a été mise en œuvre.

Cette présentation exposera en détail le mécanisme de contrôle de cohérence mis en œuvre ainsi que les estimateurs composites retenus, avant de présenter une évaluation de l'impact de ces améliorations méthodologiques sur la qualité des statistiques produites.

## Bibliographie

Brion P. (2009). L'utilisation combinée de données d'enquêtes et de données administratives pour produire les statistiques structurelles d'entreprise, article présenté aux dixièmes Journées de Méthodologie Statistique, Paris, France.

Gros E. (2012). Esane ou les malheurs de l'estimation composite : comment gérer les valeurs négatives d'estimateurs par différence ?, article présenté aux onzièmes Journées de Méthodologie Statistique, Paris, France.