

L'utilisation combinée de données d'enquêtes et de données administratives pour produire les statistiques structurelles d'entreprises

Philippe BRION¹

Le nouveau dispositif de production des statistiques structurelles d'entreprises françaises, Esane (Élaboration des Statistiques ANnuelles d'Entreprises), est 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, déclarations douanières), complétée par des données obtenues par une enquête statistique réalisée sur un échantillon d'entreprises.

La production de statistiques doit donc s'opérer à partir d'une base de données rectangulaire « incomplète » (les données administratives étant exhaustives alors que les données d'enquêtes sont obtenues sur un échantillon). Différents types d'estimateurs sont proposés pour répondre à cette demande, selon les types de statistiques à produire. En particulier, les statistiques sectorielles, qui s'appuient sur le classement des entreprises dans la nomenclature d'activités, peuvent utiliser de manière intéressante le code d'activité principale déterminé pour les unités de l'échantillon à partir de l'enquête, combiné avec le code d'activité du répertoire d'entreprises Sirene, qui n'est pas forcément à jour.

L'utilisation conjointe des deux types de sources nécessite également une phase de confrontation de données individuelles sur quelques variables, qui peut conduire à procéder à un arbitrage entre ces deux sources sur certaines unités. Elle doit être mise à profit au travers des estimateurs finaux, et permet en particulier de tenir compte d'imperfections dans l'une ou l'autre source.

¹ DSE, Insee