

Contrôles des rémunérations dans les DADS – Une analyse exploratoire pour améliorer la détection des points atypiques

Claire Jacod¹

Résumé

À l'Insee, les Déclarations Annuelles des Données Sociales (DADS) constituent la source annuelle de référence sur l'emploi et les salaires. La qualité des informations déclarées et leur adéquation aux besoins statistiques sont variables, notamment sur les rémunérations. Il faut donc les contrôler et éventuellement les corriger. Ces traitements sont actuellement réalisés au sein des applications DADS, sur le secteur privé, et Siasp (Système d'information sur les agents du service public), sur le secteur public, selon des méthodes proches, basées sur la modélisation du salaire horaire par les moindres carrés ordinaires, en utilisant des variables auxiliaires redressées.

Dans le cadre de cette étude, on propose une étape préliminaire de détection d'outliers, en amont des traitements effectués dans les applications DADS et Siasp, et donc avant le redressement des variables auxiliaires. Conformément aux attentes du futur système d'information sur l'emploi et les revenus, l'objectif est un repérage précoce et homogène sur l'ensemble du champ des valeurs atypiques, qui seront confirmées ou infirmées dans la suite du processus. Une fois ces données isolées en vue d'un traitement spécifique, il est possible de traiter sans délai les informations issues du cas nominal.

On montre que les moindres carrés ordinaires ne sont pas suffisamment robustes et qu'il est nécessaire d'utiliser une méthode d'estimation robuste non seulement aux outliers, mais également aux points leviers. On propose une méthode alternative, adaptée à la présence de variables explicatives discrètes et au volume important des données que l'on souhaite traiter.

¹ Unité Qualité - DMCSI - Insee. Au moment de l'écriture de l'article, division EFA - DERA - DSDS - Insee.

