

Mise au point de critères d'arrêt pour la chaîne de traitement EPURE

Nathalie CARON¹ et Sylvie GRCIC²

¹ Insee, Unité Méthodes Statistiques

² Insee, Département de l'Emploi et des Revenus d'Activité

La chaîne de traitement EPURE (Extension du projet URSSAF sur les revenus et l'emploi) mise en production à l'Insee en 1996 est fondée sur l'exploitation trimestrielle des bordereaux récapitulatifs de cotisations adressés aux URSSAF par les employeurs affiliés au régime général. Ces bordereaux, dont l'objet est de calculer les cotisations sociales dues par l'employeur, contiennent en particulier des informations sur le nombre de salariés inscrits au dernier jour de la période et le nombre de salariés rémunérés pendant la période, les salaires versés, les exonérations, les cotisations dues. Les données réceptionnées et traitées sont surtout utilisées pour le suivi conjoncturel des indices d'évolution d'emploi. Il existe deux niveaux de publication : un niveau national et un niveau régional.

Dans EPURE, les données sont contrôlées par le «redresseur». Il s'agit d'un ensemble de programmes informatiques qui affectent à chaque établissement une note établie principalement en fonction de la vraisemblance des évolutions constatées de l'emploi compte tenu de son passé. En cas de détection d'anomalie, le redresseur propose des corrections qui sont expertisées par les gestionnaires dans les Directions Régionales de l'Insee. Parallèlement, dans certains cas, le redresseur exclut des établissements des calculs d'indice. Les travaux manuels d'expertise et de redressement visent à inclure des établissements exclus automatiquement par le redresseur ainsi qu'à expertiser et affiner les corrections automatiques. Une vérification exhaustive des corrections apportées par le redresseur à l'ensemble des établissements repérés en anomalie est impossible ; celle-ci est donc sélective. La règle de traitement des lots d'anomalies repose actuellement sur un critère de taille. En effet, l'objectif national de publication impose aux gestionnaires de traiter en priorité les établissements de plus de 50 salariés. Lorsque cet objectif est atteint, les gestionnaires ont alors toutes les possibilités pour traiter les autres établissements en anomalie. Actuellement la fin de la phase d'expertise réalisée en direction régionale correspond au délai en termes de nombre de semaines qui leur est accordé.

L'objectif de notre démarche est de proposer un critère d'arrêt de l'expertise des gestionnaires tout en garantissant la même « qualité » dans les principaux domaines de diffusion et d'essayer de quantifier la « qualité » du fichier final obtenu. La détermination d'un critère d'arrêt de l'expertise des gestionnaires revient finalement à trouver un ordre adéquat de traitement des établissements et de fixer le seuil au-delà duquel l'expertise manuelle n'apporte rien, c'est-à-dire ne modifie plus les évolutions obtenues par domaine de diffusion ou plus exactement les modifie de manière marginale en-deça du seuil de qualité définie.

Pour réaliser cet objectif, plusieurs régions ont participé à un test sur les troisième et quatrième trimestres 2003. Le travail demandé aux équipes EPURE consistait à traiter l'ensemble des anomalies de leur région. In fine, on dispose pour chaque trimestre de trois effectifs : la valeur déclarée par l'établissement, la valeur redressée par le redresseur et la valeur corrigée par le gestionnaire. En supposant que la valeur 'vraie' est la valeur proposée par le gestionnaire, notre démarche a été de modéliser l'écart entre la valeur 'redressée' et la valeur 'vraie', de manière à être capable, par la suite, de prédire au mieux cet écart. Une fois un modèle retenu et ajusté, il faut vérifier qu'il est pertinent pour toutes les régions et quel que soit le trimestre considéré. Pour la validité de la démarche, la relation estimée doit être largement indépendante du trimestre et de la région considérés. En effet, en raison de la lourdeur du travail demandé aux directions régionales, il n'est possible de calculer la valeur «vraie» que sur un nombre de trimestres et de régions limités.