

Synthèse d'une proposition de communication aux JMS

24 octobre 2014

Titre : Calage du sous-échantillon annuel de l'enquête Emploi sur les résultats trimestriels

Auteur : Frédéric Ouradou, Insee, DMCSI, frederic.ouradou@insee.fr

Domaine : Calage

Résumé :

L'enquête emploi en continu (EEC) est une enquête trimestrielle en panel. Chaque trimestre, six vagues sont interrogées : une vague en 1^{re} interrogation, dite « entrante », la vague entrante du trimestre précédent, en 2^e interrogation... jusqu'à la vague « sortante », en 6^e interrogation. Certaines questions de l'enquête ne sont posées qu'en vague 1, d'autres seulement en vagues 1 et 6. Pour les variables correspondantes, les statistiques annuelles sont donc fondées sur un sous-échantillon de l'EEC, constitué de la compilation sur l'année des vagues pour lesquelles on dispose de ces variables. Un règlement d'Eurostat relatif à l'enquête Emploi impose, dans ce cas, qu'une « *cohérence entre les totaux annuels des sous-échantillons et les moyennes annuelles des échantillons complets (soit) assurée pour l'emploi, le chômage et l'inactivité par sexe et pour les tranches d'âge suivantes : 15-24 ans, 25-34 ans, 35-44 ans, 45-54 ans et 55 ans ou plus* ». La contribution expose la méthode utilisée pour garantir cette cohérence.

Pour le calage des totaux du « sous-échantillon annuel » sur les moyennes annuelles des totaux trimestriels, dit plus simplement calage annuel, on peut souhaiter utiliser des marges de calage supplémentaires (niveau individu ou ménage), en plus de celles requises par Eurostat. Les marges peuvent être des marges déjà utilisées en trimestriel pour l'EEC, ou bien des marges déduites des résultats trimestriels de l'EEC. La cohérence du calage peut être annuelle ou bien trimestrielle, et dans ce dernier cas, les totaux correspondant à la vague 1, ou aux vagues 1 et 6 d'un trimestre donné doivent correspondre aux totaux du trimestre sur l'ensemble des six vagues.

Un certain nombre de scénarios de calage sont réalisés, ces scénarios différant selon les variables de calage, voire selon les modalités des variables de calage, selon la méthode de calage, *Raking ratio* ou *Logit*, selon la taille de l'échantillon à caler (compilation des vagues 1 trimestrielles ou bien des vagues 1 et 6), et selon le type de cohérence requise pour le calage (cohérence trimestrielle ou annuelle).

Les différents scénarios sont comparés selon plusieurs critères : nombre d'itérations du logiciel Calmar pour arriver à la convergence, statistiques de distribution des poids finals, statistiques de distribution des rapports de poids, troncature des poids extrêmes, effet du calage sur certaines variables dites « de contrôle ».

On testera notamment la robustesse du calage vis-à-vis du nombre de variables de calage, de la fenêtre des rapports de poids dans le cas de la méthode Logit, et on s'intéressera au comportement de Calmar en présence d'observations présentant des valeurs manquantes pour certaines variables de calage.