

Prise en compte des équilibrages dans la précision transversale de l'enquête Emploi

Karim MOUSSALLAM¹

L'étude compare plusieurs méthodes d'estimation de la variance transversale de l'enquête Emploi sur la base de simulations du tirage. Son champ couvre la partie de l'estimateur validable par ces simulations, donc hors prise en compte de la non-réponse.

La première partie de l'étude 'revisite' la question de la pondération optimale des résidus d'équilibrage pour le calcul de variance, sur la première phase de l'enquête Emploi. L'article de Deville et Tillé sur ce sujet est appliqué dans le contexte particulier d'un nombre élevé de variables d'équilibrage et d'une taille d'échantillon d'unités primaires faible pour certaines régions. La disponibilité des bases de sondage des grappes de secteurs et de logements issues de la Taxe d'Habitation (TH) permet de comparer les résultats des estimateurs de variance envisagés aux écarts quadratiques moyens observés sur un nombre élevé d'échantillons, simulés par le plan de sondage effectivement utilisé.

La méthode d'estimation la plus satisfaisante théoriquement nécessite un algorithme de résolution d'une équation non linéaire. Après un changement de variable pour s'abstraire des contraintes de bornes admissibles, une méthode de 'descente du gradient' a paru préférable pour rechercher une solution. Mais la mise en œuvre soulève des difficultés techniques. Le calcul du gradient est affecté par l'instabilité numérique de l'inverse généralisée, c'est-à-dire que les conditions de Moore et Penrose définissant celle-ci ne sont pas exactement vérifiées. Le grand nombre de variables d'équilibrage pour le tirage de première phase semble exclure de pouvoir vérifier exactement l'équation déterminante.

Les résultats confirment assez clairement que la méthode la plus simple pour pondérer les résidus de l'équilibrage a davantage tendance à sous-estimer la variance que les alternatives envisagées. A cet égard, la pondération 'implicite' paraît la meilleure. Cependant, cet 'optimum' s'avère moins satisfaisant que dans l'étude de Deville et Tillé, parce qu'il surestime plus sensiblement la variance de certaines des variables testées. Ceci pourrait être dû soit à la difficulté d'optimiser la pondération, soit à ce que le cadre théorique de l'approximation de variance n'est pas bien adapté au plan de sondage effectif, qui serait donc trop éloigné d'un 'poissonien équilibré'.

La seconde partie cherche à répondre à la question de la prise en compte de l'équilibrage dans le contexte d'un plan de sondage où une seule unité est échantillonnée par 'strate', et un équilibrage avec deux étapes d'atterrissage inter-strates. Pour cela, deux méthodes d'estimation de la variance de deuxième phase de l'enquête EEC-TH sont comparées. L'approche la plus simple assimile le tirage intra-secteur à un degré d'échantillonnage. Mais des simulations complémentaires du tirage de deuxième phase suggèrent qu'il serait préférable de prendre en compte l'équilibrage effectué à ce niveau également par la technique des résidus, malgré le prix de l'omission de certaines covariances. La simplification du plan de sondage pour le calcul de variance fait le pendant du collapse des unités primaires requis par la première méthode.

¹ karim.moussallam@insee.fr, Insee DSDS UMS

Au total, malgré l'ampleur de l'incertitude sur la variance de première phase et dans les limites notamment du faible nombre de variables examinées, l'estimateur de la variance d'échantillonnage de l'enquête Emploi paraît acceptable au vu des résultats des simulations effectuées.