

Intégration des enquêtes sur les ménages pour la collecte de données financières à l'ONS

Salah Merad et Fola Ariyibi

Office for National Statistics, UK

Salah.merad@ons.gsi.gov.uk

Résumé

A l'ONS actuellement nous conduisons plusieurs enquêtes de ménages qui recueillent des données sur le revenu, en plus des variables spécifiques à chaque enquête. Nous travaillons à leur intégration progressive dans une seule enquête avec une conception modulaire pour améliorer la cohérence et la précision des estimations des variables communes. L'une de ces enquêtes est longitudinale et possède un plan d'échantillonnage à rotation de quatre vagues que nous étendrons sur six vagues.

Dans cette présentation, nous allons décrire la conception du plan d'échantillonnage de l'enquête intégrée, évaluer les améliorations attendues dans la précision des estimations et discuter de la manière dont nous allons gérer la transition de l'ensemble actuel d'enquêtes à la nouvelle enquête intégrée.

Introduction

A l'Office for National Statistics du Royaume-Uni, plusieurs enquêtes de ménages recueillent des données sur le revenu. Les plus importantes sont : l'enquête de l'UE sur le revenu et le coût de la vie (EU SILC), l'enquête sur le coût de la vie et la nourriture (LCF), et l'enquête de la richesse des ménages et de l'actif (WAS). Afin de réduire les coûts d'enquête et de produire des estimations du revenu qui sont cohérentes et avec une plus grande précision, le Bureau a commencé à travailler sur l'intégration graduelle de ces enquêtes en une seule enquête avec plusieurs modules.

Plan d'intégration

Dans la première étape, nous cherchons à intégrer les enquêtes EU-SILC et LCF. Cette dernière est une enquête transversale qui recueille des données sur le revenu des ménages et des données détaillées sur les dépenses sur la nourriture et d'autres articles à travers l'utilisation d'un journal intime. Les données sur le revenu tirées de cette enquête sont utilisées dans la publication d'une importante étude sur les effets de la fiscalité et des prestations sociales. L'enquête EU-SILC a un plan longitudinal à quatre vagues qui vise à recueillir des données détaillées sur le revenu; les statistiques obtenues de cette enquête sont requises de satisfaire la réglementation spécifiée par l'UE. Actuellement, la première vague de cette enquête est un sous-échantillon d'une enquête qui est gérée par le ministère du Travail et des pensions; il a donc été difficile de satisfaire certaines règles de l'UE sur la couverture de la population et la définition de certaines variables.

Dans l'avenir nous envisageons d'avoir une enquête intégrée avec une conception modulaire; dans la première étape, il y aura un module pour les variables clés du revenu, un module de données détaillées du revenu utilisées pour l'EU-SILC, et un module sur les données sur les dépenses. Dans le cadre de la première vague, l'ensemble de l'échantillon seront priés de remplir le module contenant les principales variables sur le revenu. L'échantillon sera divisé en deux sous-échantillons non chevauchants, et chaque ménage sera invité à remplir l'un de l'EU SILC ou LCF modules. Les ménages dans le sous-échantillon recevant le module UE SILC seront encore contactés dans les 5 prochaines années, et formeront l'ensemble de données longitudinales. L'enquête a un design en deux étapes où les secteurs postaux constituent les unités primaires d'échantillonnage; pour la précision, il serait idéal d'avoir les deux modules dans chaque secteur de code postal mais cela pourrait causer des difficultés aux intervieweurs. Nous considérons les façons d'aborder les risques par l'amélioration de la gestion du processus de collecte et de la formation des intervieweurs .

Nous effectuons une étude sur la taille de l'échantillon requise pour chaque module, avec un accent particulier sur la satisfaction et les règlements de l'UE sur la précision des indicateurs de pauvreté. L'une des exigences est liée aux estimations régionales au niveau NUTS2; nous considérons si la stratification de l'échantillon par NUTS2, au lieu de NUTS1 métropolitaine/non métropolitaine, comme cela se fait actuellement, va améliorer la précision au niveau régional et au niveau national. Nous examinons également les répercussions d'un changement dans la stratification sur les estimations de dépenses. En plus de stratifier par la région, nous utilisons également des stratificateurs basés sur des facteurs socio-démographiques obtenus du recensement général de la population, par exemple la distribution de l'âge de la population et la prévalence du chômage et des ménages vivant dans des logements sociaux. Ces facteurs sont transformés en variables catégoriques pour tenir compte de la probabilité proportionnelle à la taille de l'échantillonnage systématique à partir d'une trame triée de secteurs postaux qui est utilisée pour sélectionner les unités d'échantillonnage primaires. Parce que seul un nombre relativement restreint de secteurs postaux est sélectionné dans chaque région, le nombre de facteurs qui peuvent être utilisés pour trier la trame est limitée; nous déterminons le meilleur ensemble de facteurs à utiliser pour les deux enquêtes EU SILC et LCF. L'évaluation des stratificateurs est effectuée à l'aide de modélisation de régression des variables d'enquête où les stratificateurs sont les variables indépendantes.

Pour déterminer la taille de l'échantillon requise pour l'enquête EU-SILC, nous estimons les effets de conception des estimations au niveau NUTS2 pour les deux méthodes de partitionnement de l'échantillon en deux modules pour LCF et EU SILC. Nous utilisons les données actuelles pour estimer les coefficients d'intraclasses et ensuite utiliser la méthode d'approximation standard pour produire les estimations des effets de conception. Pour les indicateurs de pauvreté de EU-SILC, qui sont des fonctionnelles de quantiles, nous utilisons la méthode de linéarisation pour l'estimation de la variance proposée par Deville (Deville, 1999). Cette méthode nécessite l'élaboration d'une fonction d'influence; une formule est disponible pour l'indicateur 'taux de la population menacée de pauvreté', mais elle n'est pas disponible pour d'autres indicateurs tels que "à risque de pauvreté ou d'exclusion sociale", qui est une combinaison de trois indicateurs de pauvreté. La méthode de linéarisation de Taylor semble surestimer les effets de conception pour l'indicateur "à risque de pauvreté", ce qui signifie qu'il peut ne pas être fiable à utiliser. Nous discutons des moyens d'obtenir des estimations des effets de conception plus fiables pour l'indicateur "à risque de pauvreté ou d'exclusion sociale".

Ajustement for la non-reponse et ponderation

L'ajustement en cas de non-réponse est fait différemment dans les deux enquêtes : dans LCF nous utilisons des facteurs de réponse dérivés de l'étude qui lie les données du recensement de la population et les données d'enquête recueillies autour du moment du recensement; dans l'enquête EU SILC la non-réponse est traitée lors de la phase de calage. Nous examinons l'utilité des données publiquement disponibles pour ajuster pour la non-réponse à travers une méthode commune qui est la pondération de l'échantillon, et d'évaluons l'impact sur les estimations des variables clés.

Dans l'enquête UE SILC l'ensemble de données longitudinales passera de quatre vagues à six vagues, ce qui présentera un défi pour le suivi des personnes qui quittent les ménages dans l'échantillon original. Nous examinons nos procédures de suivi actuelles et considérons des simplifications pour réduire les coûts du suivi sans nuire à la qualité des données. Un plan longitudinale à six vagues entraînera un plus gros échantillon pour produire l'ensemble de données pseudo transversales; cependant, les six panneaux seront sélectionnés à partir de populations couvrant une grande période de temps, ce qui entraîne une couverture différentielle entre les panneaux. Nous examinons comment cela affecte les façons dont les panneaux sont combinés pour produire les poids pseudo transversaux.

Dans la pondération des modules de l'EU SILC et LCF nous introduisons des contrôles de calage pour les variables clés du revenu, ce qui conduira à des résultats uniformes dans les modules. Nous décrivons le système de pondération pour produire des poids transversaux pour les différents modules, les poids longitudinaux de l'EU SILC pour toutes les vagues, ainsi que les poids pseudo transversaux

Challenges de la transition

Pour un certain nombre d'années, nous allons traiter les données recueillies de la présente série d'enquêtes et les données recueillies dans l'enquête intégrée. Ceci affectera en particulier l'ensemble des données longitudinales de l'UE SILC; nous examinons les façons de minimiser le manque de cohérence entre les anciens et les nouveaux panneaux.

References

Deville, J. C. (1999). Estimation de la variance pour des statistiques complexes: linearization et techniques des résidues. *Survey Methodology*, 25, 193-203.