

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

Salah Merad et Fola Ariyibi

ONS

Colloquium Francophone de méthodes d'enquêtes
Gatineau, 12-14 Octobre 2016

Plan de présentation

- Vision pour les enquêtes sur les finances des ménages
- Objectifs et contraintes
- Méthode d'intégration des plans d'échantillonnage
- Travaux futurs

Vision

- Première étape: intégration de SILC et LCF
 - SILC: longitudinale; données détaillées sur le revenu (indices of pauvreté)
 - LCF: transversale; données sur le revenu des ménages et des données détaillées sur les dépenses (journal alimentaire)
- Deuxième étape: intégrer l'enquête sur la richesse des ménages et de l'actif - panel avec des boosts réguliers

SILC

- Passer d'un plan avec 4 vagues à un plan avec 6 vagues
- les statistiques obtenues de cette enquête sont requises de satisfaire la réglementation spécifiée par l'Eurostat: seuils de précision au niveau de geograpghy "NUTS2"(39 régions)
 - La taille effective de l'échantillon est une fonction du nombre de ménages dans la région

$$n_{eff,h} = a\sqrt{N_h} + b$$

Variables communes

- LCF et SILC ont plusieurs variables en commun
- Les variables sur le revenu sont communes et donc pourraient être utilisées dans l'estimation des indicateurs de pauvreté: utiliser des questions identiques dans les deux enquêtes
- Réponses collectées des deux modules seront utilisées pour satisfaire le règlement de l'Eurostat sur la précision des estimations des indices de pauvreté

Plan d'intégration: objectifs et contraintes

- Maximiser l'amélioration de la précision des estimations des variables communes
- Minimiser l'impact sur les estimations de dépenses obtenues de LCF
- Satisfaire les demandes en précision spécifiées par l'Eurostat
- Operationnellement pratique sur le terrain

Plan d'échantillonnage

- Echantillon à deux étapes dans les deux enquêtes: unités primaires sont des secteurs postaux
- Stratification majeure: NUTS2
 - Les secteurs sont ordonnés par deux facteurs socio-démographiques obtenus du recensement général de la population

Impact du changement de stratification sur LCF

- Régression des dépenses moyennes par secteur sur la region and des caractéristiques obtenues du recensement

Model	Adjusted R-square		
	2012	2013	2014
GOR4, HNSSEC3, HNOCAR	0.34	0.34	0.34
NUTS2, HNSSEC3, HNOCAR	0.35	0.35	0.34

Plans d'intégration

Design	Propriétés
Deux modules par secteur 12 cas pour LCF et 10 cas pour SILC	Effet de regroupement plus bas pour les deux modules Effet de regroupement plus haut pour les questions communes Moins efficace sur le terrain Plus grand challenge pour les intervieweurs
Un module par secteur 18 cas pour LCF 15 cas pour SILC	Effet de regroupement plus haut pour les deux modules Effet de regroupement plus bas pour les questions communes Plus pratique pour les intervieweurs

Taille requise de l'échantillon entier

- Pour la strate h ,

$$n_{r,h}^* \geq Deff_h n_{eff,h}$$

$n_{eff,h}$ est la taille effective spécifiée par Eurostat

$$Deff_h \cong 1 + (\bar{m}_{r,h} - 1) \rho_h$$

$\bar{m}_{r,h}$ est le nombre moyen d'observations dans une unité primaire

ρ_h est le coefficient intraclasse

Calculution de l'effet du plan

- Paramètre: Proportion de personnes dont le revenu du ménage est en dessous de 60% du revenu médian
- Méthode de calcul de la variance: linearisation de Deville
 - Suppose que le revenu médian est estimé
 - Basée sur une fonction d'influence
- Utilisé des données des 3 dernières années
- Déterminé la valeur du coefficient intraclasse pour chaque année: pris la valeur moyenne
- Calculé l'effet du plan basé sur 8 réponses par secteur (chiffres de la première vague)

Estimation des erreurs d'échantillonnage

NUTS 2	Standard	Linearisation de Deville
UNITED KINGDOM	0.70	0.54
Tees Valley and Durham	5.45	5.50
Cumbria	5.71	5.74
Greater Manchester	3.57	3.55
Lancashire	3.58	3.62
Cheshire	4.65	4.68

Méthode de calcul du nombre nécessaire d'unités primaires

- Basée sur l'indice "A risk de pauvreté"
- Données utilisées pour l'estimation de cet indice proviendront de:
 - Module LCF
 - Module SILC, vagues 1 à 6
- Un module par secteur
- Module LCF: nombre d'unités primaires égal au nombre actuel; allocation proportionnelle

Allocation pour le module SILC (1)

- Nombre moyen de ménages qui répondent au module LCF

$$n_{r,LCF,h} = n_{I,LCF,h} \bar{m}_{r,LCF,h}$$

- Nombre requi de réponses SILC sur 6 vagues

$$n_{r,SILC,h}^* = Deff_{\max,h} n_{eff,h} - n_{r,LCF,h}$$

- Nombre d'unités primaires pour le module SILC

$$n_{I,SILC,h}^* = n_{r,SILC,h}^* / \bar{m}_{r,v1-v6,h}$$

Allocation pour le module SILC (2)

- Période de transition des plans actuels au nouveau plan
 - Réaliser des niveaux de précision similaires dans toutes les regions NUTS2
 - Augmenter le nombre d'unités primaires de telle façon que
Erreur d'échantillonnage soit moins de 50% plus élevée que le seuil requi par l'Eurostat

Practicalités dans la collection

- Un bon nombre de secteurs ne contiennent que quelques cas de SILC après la 3ème vague: inefficace sur le terrain
- Nous considérons retenir la moitié des secteurs pour une autre année
 - Augmentation du nombre de ménages par secteur: collection devient efficace
 - Covariance entre les vagues: augmentation de l'effet du plan

Travaux pour le futur

- Intégration de l'enquête sur la richesse des ménages et de l'actif (collection de données sur deux ans); plan longitudinal
- Utilisation de données administratives
 - Dans le design de l'échantillon
 - Pour remplacer certaines questions
 - Ajustement pour la non-réponse et l'estimation