

ANALYSE COMPARATIVE DES METHODES D'IMPUTATION DES DONNEES MANQUANTES SUR LES REVENUS MESURES DANS LES ENQUETES AUPRES DES MENAGES SUR L'EMPLOI DANS LES PAYS D'AFRIQUE SUBSAHARIENNE : CAS DU CAMEROUN, DU CONGO ET DE LA COTE D'IVOIRE

Ousman KORIKO

Ingénieur Statisticien Démographe

Observatoire Economique et Statistique d'Afrique Subsaharienne (AFRISTAT)

BP E 1600, Tél. 0022376452488, Fax 0022320211140 Bamako, Mali,

ousman.koriko@afriostat.org

Résumé :

Le présent article s'intéresse essentiellement au traitement des données manquantes sur le revenu du travail en étudiant les cas constatés dans les enquêtes réalisées au Cameroun en 2005, au Congo en 2009 et en Côte d'Ivoire en 2012 avec les mêmes outils proposés par AFRISTAT. Selon les résultats de ces enquêtes, plus de la moitié des travailleurs qui sont rémunérés déclarent leur revenu de façon détaillée. Par contre, un nombre relativement important de travailleurs préfèrent déclarer leur revenu en tranche. Tandis que d'autres ne déclarent pas leur revenu. Dans ces deux derniers cas, l'article propose des méthodes d'imputation. L'article décrit la qualité des données collectées sur les revenus liés au travail dans les enquêtes auprès des ménages sur l'emploi dans les pays d'Afrique subsaharienne et présente des imputations de données manquantes selon des approches déterministes et non déterministes. Les résultats obtenus sont analysés en vue d'une orientation en matière de traitement des données manquantes sur le revenu ou d'autres agrégats économiques.

Mots clés : revenu, emploi, imputation, non réponse, Cameroun, Congo, Côte d'Ivoire

Introduction

L'Observatoire Economique et Statistique d'Afrique Subsaharienne (AFRISTAT), a développé au profit de ses Etats membres des outils harmonisés pour la collecte, le traitement et l'analyse des données des enquêtes auprès des ménages sur l'emploi. Ces enquêtes ont pour objectifs spécifiques entre autres : i) de fournir les principaux indicateurs du marché du travail, ii) d'étudier l'évolution du marché du travail, notamment en terme d'offre de main d'œuvre de la part des ménages, iii) d'analyser l'évolution des conditions d'activité et les déterminants de la mobilité de l'emploi et iv) de déterminer les revenus du travail et d'analyser leur distribution.

Le présent article s'intéresse essentiellement au traitement des données manquantes sur le revenu du travail, en étudiant les cas constatés dans les enquêtes réalisées au Cameroun en 2005, au Congo en 2009 et en Côte d'Ivoire en 2012 avec les mêmes outils proposés par AFRISTAT.

1. Qualité des données sur les revenus du travail

Il est important de souligner d'entrée que l'enquête du Congo en 2009 a couvert uniquement les villes de Brazzaville et de Pointe Noire qui cumulent plus 63% de la population totale du

pays. Tandis que pour les deux autres pays, il s'agit des enquêtes dont la couverture est nationale. Au sens des trois enquêtes, est pourvue d'un emploi, toute personne en âge de travailler, ayant exercé au cours d'une période de sept jours précédant le passage de l'enquêteur dans le ménage où elle réside, une activité productive au sens du système de comptabilité nationale (SCN) contre une rémunération. Dans le cadre du présent article et pour des raisons de comparaison, nous considérons la tranche d'âge 15-64 ans comme étant l'âge de travailler. Seuls les revenus tirés des activités principales seront analysés ici.

L'analyse des résultats des trois enquêtes montre que les types de revenus du travail ne sont pas les mêmes d'un pays à l'autre. Toutefois, la Côte d'Ivoire et le Cameroun semblent être plus proches dans leur structure de paiement des travailleurs, dominée essentiellement par le secteur privé où les revenus basés sur les bénéficiaires sont plus fréquents. Par contre au Congo, le secteur public est encore dominant avec le paiement de salaire fixe.

On relève cependant que 24,1%, 15,5% et 0,7% des travailleurs respectivement ivoiriens, camerounais et congolais qui sont rémunérés, n'ont pas déclaré leur revenu. Les données ont permis de constater que les travailleurs salariés sont généralement moins réticents à déclarer leur revenu. Par ailleurs, en analysant les données des échantillons enquêtés, la distribution du revenu déclaré par les travailleurs semble plus homogène au Congo qu'au Cameroun et surtout en Côte d'Ivoire. Les coefficients de variation se chiffrent respectivement à 1,32, 1,54, 1,95 et témoignent dans tous les cas, du caractère disparate des niveaux de revenus.

2. Approches d'imputation

Les documents de cours de Sébastien Le Faivre (2013), de Guillaume Chauvet (2012) et le livre de Pascal Ardilly (1994) abordent suffisamment les méthodes de traitement des non réponses partielles. Dans cet article, une approche déterministe et une approche aléatoire sont utilisées pour les imputations des revenus. S'agissant d'abord des revenus déclarés en tranche, nous avons procédé à une estimation des valeurs exactes bornées par les valeurs de l'intervalle de revenu correspondant. Dans la méthode déterministe, les centres des classes sont utilisés comme des revenus moyens attribués aux travailleurs concernés. Dans l'approche non déterministe, il a été généré pour chaque catégorie socioprofessionnelle, un nombre aléatoire borné sur l'intervalle, selon la loi uniforme. L'hypothèse fondamentale est que le revenu évolue de façon uniforme selon les catégories socioprofessionnelles des travailleurs dont la nomenclature est présentée ci-après : 1) Cadre supérieur, ingénieur et assimilé, 2) Cadre moyen, agent de maîtrise, 3) Employé/ouvrier qualifié, 4) Employé/ouvrier semi qualifié, 5) Manœuvre, 6) Employeur, 7) Travailleur pour propre compte, 8) Autre catégorie.

Concernant les travailleurs rémunérés n'ayant pas déclaré de revenu, l'imputation des valeurs est faite par la méthode de régression log linéaire. Les variables dépendantes pour calculer le modèle, sont les revenus exacts déclarés et les revenus exacts estimés pour les travailleurs ayant déclaré leur rémunération en tranche. Les variables explicatives du modèle sont travailleurs (âge, sexe, niveau d'instruction, milieu de résidence, région) ainsi que de leurs attributs dans le domaine de l'emploi (métier, catégorie socio professionnelle, branche d'activité et secteur institutionnel).

Pour l'approche déterministe, seuls les coefficients de régression du modèle sont récupérés et appliqués pour l'imputation du revenu des travailleurs qui ne l'ont pas déclaré. Par contre dans le cas de l'approche non déterministe, en plus des coefficients de régression, les résidus du modèle sont récupérés. Les résidus sont distribués de façon aléatoire selon une loi de probabilité uniforme bornée dans les fourchettes des valeurs résiduelles fournies par le modèle de régression.

3. Précision des estimations

Il est calculé selon le plan de sondage de chaque pays, l'estimateur du revenu moyen et le coefficient de variation de l'estimateur dans les trois scénarii suivants : i) estimateur basé uniquement sur les personnes ayant déclaré leur revenu en détail, ii) estimateur calculé après imputation des revenus détaillés pour les personnes ayant déclaré des revenus en tranche, iii) estimateur calculé après imputation des revenus détaillés pour les personnes n'ayant déclaré des revenus ni en détail ni en tranche. Les estimateurs des variances des estimateurs ont été calculés selon la méthode analytique.

4. Analyse des résultats

L'analyse comparative est faite en deux temps, d'abord en intra puis en inter pays. Au Congo, où l'enquête a couvert exclusivement les deux principales villes du pays, très peu de travailleurs rémunérés n'ont pas fourni d'information sur leur revenu. Par ailleurs, le fait que la majorité des travailleurs perçoit un salaire fixe, cela a permis d'obtenir des informations plus cohérentes. Le revenu moyen de l'ensemble des travailleurs est estimé à 102 750 FCFA à partir des travailleurs qui ont uniquement déclaré leur revenu en détail, avec un coefficient de variation de 9,3%. Une amélioration du coefficient de variation est obtenue à la suite de l'imputation des revenus détaillés pour les personnes ayant déclaré leur gain en tranche. Le coefficient de variation se situe à 9% selon la méthode déterministe et 9,1% avec la non déterministe. A la suite de l'imputation des revenus pour les personnes qui n'ont pas fait de déclaration, le coefficient de variation a baissé de nouveau. La méthode non déterministe l'a ramené à 8,5%, soit une amélioration de 0,8 point.

Au Cameroun, avant toute imputation, le revenu moyen des travailleurs est estimé à 35 180 FCFA, avec un coefficient de variation de 6%. L'estimation à la suite de l'imputation par la méthode non déterministe a fait passer le revenu moyen à 41 480 FCFA et 44 260 FCFA respectivement pour les scénarii 2 et 3. Les coefficients de variation sont passés respectivement à 5,6% et 5,2% dans les deux cas. Il y a ici aussi une amélioration de la précision des estimations de revenus à la suite des imputations, notamment par l'approche probabiliste ou non déterministe.

En Côte d'Ivoire, dans le premier scénario, le revenu moyen estimé s'élève à 58 370 FCFA, avec un coefficient de variation de 7,3%. Le revenu moyen estimé a atteint 64 180 FCFA dans le second scénario et 70 550 FCFA dans le troisième, après imputation par la méthode non déterministe. Les coefficients de variation sont aussi améliorés en baissant respectivement à 6% et 5,1%.

Il résulte de cette analyse, quel que soit le pays, que l'imputation des revenus par la méthode non déterministe améliore la précision des estimations. Cette précision dépend aussi de la qualité des données déclarées et du plan de sondage utilisé. Dans les trois pays, un plan de sondage stratifié à deux degrés, avec probabilité inégale au premier degré a été utilisé. Par ailleurs, les tailles d'échantillon sont suffisamment importantes pour permettre des estimations significatives.

S'agissant de la comparaison entre les pays, les imputations par la méthode non déterministe ont plus amélioré les données sur les revenus des travailleurs de la Côte d'Ivoire, avec une baisse du coefficient de variation de l'estimateur du revenu de 2,2 points. Par contre pour le Cameroun et le Congo, la baisse constatée est de 0,8 point. La distorsion des données de base avant les imputations semble alors plus sévère au Cameroun et au Congo. Par ailleurs, il est aussi important de relever que dans le deuxième scénario, quel que soit le pays, l'imputation par la méthode déterministe améliore plus que la méthode non déterministe, le coefficient de variation de l'estimateur du revenu moyen. Il est prudent dans ce cas de définir pendant l'enquête des tranches de revenu moins larges pour éviter d'obtenir des centres de classe trop élevés. Cette conclusion s'inverse dans le scénario 3, à la suite de l'imputation des revenus

pour les individus qui n'ont pas fait de déclaration. Dans ce cas, c'est l'approche non déterministe qui améliore la précision des estimations.

Conclusion

Pour conclure, le choix de la méthode d'imputation des données économiques manquantes (ici il s'agit du revenu du travail) dépend d'une part de la qualité des données déclarées mais aussi de la structure de l'économie nationale. Dans le cas où les données déclarées sont nombreuses, cohérentes et stables, l'imputation des données manquantes par la méthode déterministe ou non améliore la qualité des estimations.

S'agissant de l'imputation de revenu détaillé pour les individus ayant déclaré leur gain en tranche, elle améliore beaucoup plus la précision des estimations par l'approche déterministe. Par contre, le recours à la méthode non déterministe est plus indiqué pour l'imputation des revenus des travailleurs qui n'ont pas fait de déclaration.

Bibliographie

- [1] Sébastien Le Faivre (2013), Les méthodes d'imputation, document de travail, INSEE, Paris, France
- [2] Guillaume Chauvet (2012), données manquantes dans les enquêtes, cours dispensé à l'ENSAI, Rennes, France
- [3] Pascal Ardilly (1994), les techniques de sondage, éditions TECHNIP, Paris, France