

# **CALCUL DE LA PRÉCISION DES ESTIMATIONS LONGITUDINALES DANS L'ENQUÊTE SUISSE SUR LA VALEUR AJOUTÉE**

Lionel QUALITÉ<sup>1</sup>, Yves TILLÉ<sup>2</sup>

## **RÉSUMÉ**

Nous donnons une estimation de la précision des évolutions de certaines grandeurs calculées sur un panel d'entreprises subissant un taux de rotation involontaire de l'ordre de 10 à 15 %, dû à la non-réponse. Nous faisons l'hypothèse d'une non-réponse ignorable dans les strates, et le calcul tient compte des différents aspects de l'enquête : stratification, post-stratification, calage et robustification. Nous proposons ensuite des estimateurs pour les évolutions plus précis que ceux utilisés actuellement.

## **1. INTRODUCTION**

### **1.1 L'enquête WS**

En Suisse, L'office fédéral de la statistique (OFS) réalise chaque année une enquête sur la valeur ajoutée des entreprises. Il s'agit d'un panel d'établissements, stratifié par secteur d'activités et par taille d'entreprise, remis à jour périodiquement. L'échantillon effectivement observé présente un taux de rotation involontaire compris entre 10 et 15 %, entièrement dû à la non-réponse. Entre 1999 et 2001, il y a eu chaque année environ 6 000 répondants sur un échantillon de départ de 11 210 entreprises. Pour améliorer les estimations transversales, plusieurs techniques sont mises en œuvre par l'OFS : création de post-strates de « très grosses entreprises » traitées séparément ; post-stratification, calage et robustification pour les autres entreprises.

### **1.2 Objet**

Notre travail a pour but d'estimer correctement la précision des estimations longitudinales « naturelles » : les évolutions calculées à partir des estimations transversales, en tenant compte des différents aspects de l'enquête. Le taux de sondage dans certaines strates est élevé et on ne peut pas négliger les corrections de populations finies comme il est proposé dans Breidt et Fuller (1999) et Caron et Ravalet (2000).

## **2. TRAVAIL RÉALISÉ**

### **2.1 Méthode utilisée**

Une partie des techniques utilisées pour ces estimations transversales peut être prise en compte par des techniques de résidus (voir Deville, 1999). Il reste cependant le problème des changements de post-strates. Ceux-ci ne résultent que rarement de changements d'activités, mais ils proviennent surtout de la définition de la post-strate « surprise » (voir Hidioglou et Srinath, 1981) de très grandes entreprises. Elle est en effet

---

<sup>1</sup> Lionel Qualité, université de Neuchâtel, place de l'Europe 4, 2000 Neuchâtel, Suisse, e-mail : lionel.qualite@unine.ch.

<sup>2</sup> Yves Tillé, université de Neuchâtel, e-mail : yves.tille@unine.ch

définie comme regroupant les 5 % de plus grandes entreprises observées dans l'échantillon chaque année, et présente donc un assez fort taux de rotation dû aux changements de taille des entreprises, à la non-réponse, et à la limite fixée chaque année pour appartenir à cette strate, qui dépend de l'échantillon observé.

La post-strate de très grandes entreprises est considérée par l'OFS comme étant recensée dans les estimations transversales, et chaque entreprise y appartenant se voit affecter un poids égal à un. Dans le même état d'esprit, nous avons isolé les entreprises qui étaient classées dans cette strate aux deux vagues consécutives et nous les avons écartées lors du calcul de variance.

Nous avons considéré que la non-réponse était ignorable dans chacune des strates de tirage auxquelles on a soustrait les très grandes entreprises. Conditionnellement aux tailles d'échantillons obtenues, on est donc dans chaque strate en présence d'un plan de sondage simple bidimensionnel (voir Goga, 2003).

Si, dans une strate de taille  $N$ ,  $s_1$  est l'échantillon à la date  $t$  et  $s_2$  à la date  $t+1$ , si  $n_c = \text{Card}(s_1 \cap s_2)$ ,  $n_a = \text{Card}(s_1 - s_2)$  et  $n_b = \text{Card}(s_2 - s_1)$ , conditionnellement à  $(n_a, n_b, n_c)$ ,

$$P(s_1, s_2) = \frac{n_a! n_b! n_c! (N - n_a - n_b - n_c)!}{N}$$

Sous ces hypothèses, la variance de l'estimateur de Horvitz-Thompson  $\hat{Y}_t$  du total d'une variable  $Y$  à la date  $t$  (resp.  $t+1$ ) se calcule en utilisant la formule bien connue pour un plan simple de taille  $n_a + n_c$  (resp.  $n_b + n_c$ ).

Et pour calculer la variance de l'estimateur d'évolution,

$$V(\hat{Y}_{t+1} - \hat{Y}_t) = V(\hat{Y}_{t+1}) + V(\hat{Y}_t) - 2\text{Cov}(\hat{Y}_{t+1}, \hat{Y}_t),$$

il suffit de constater que

$$\text{Cov}(\hat{Y}_{t+1}, \hat{Y}_t) = N^2 \left( \frac{n_c}{n_a n_b} - \frac{1}{N} \right) S_{y_{t+1}, y_t}.$$

On peut également calculer la précision de l'estimateur de l'évolution relative en utilisant une technique de linéarisation.

## 2.2 Perspectives ultérieures

Les estimateurs utilisés, même s'ils sont relativement précis grâce à l'utilisation d'un panel, ne sont pas les meilleurs pour estimer des évolutions. Le simple fait de n'utiliser que les unités communes à deux vagues successives, comme proposé dans Caron et Ravalet (2000) permettrait en effet de diviser la variance des estimations par deux, et on peut encore réaliser des gains supplémentaires en cherchant des combinaisons linéaires optimales d'estimateurs élémentaires. Le calage et la robustification ont été réalisés dans un souci de précision transversale, et la prise en compte du caractère longitudinal de l'enquête pourrait permettre d'améliorer les estimateurs.

## RÉFÉRENCES

Breidt, F.J. et Fuller, W.A. (1999), « Estimation for supplemented panels », *Sankhyā : The Indian Journal of Statistics, special issue on sampling surveys*, Volume 61, series B, p. 58-70.

Caron, N. et Ravalet, P. (2000), « Estimation dans les enquêtes répétées : application à l'enquête emploi en continu », Rapport technique N°0005, Méthodologie Statistique, Paris, INSEE.

Deville, J.C. (1999), « Variance estimation for complex statistics and estimators : linearization and residual techniques », *Techniques d'enquêtes, survey methodology*, 25, p. 193-204.

Goga, C. (2003), *Estimation de la variance dans les sondages à plusieurs échantillons et prise en compte de l'information auxiliaire par des modèles non paramétriques*. Thèse de doctorat, Université de Rennes II, Haute Bretagne, France.

Hidioglou, M.A. et Srinath, K.P. (1981), « Some estimators of a population total from simple random samples containing large units », *Journal of the American Statistical Association*, 76, p. 690-695.