

L'impact de différentes stratégies du traitement des non-réponses dans les enquêtes auprès des entreprises

Ph. Brion
Insee





Introduction

- Enquêtes auprès des entreprises réalisées par courrier → problème des relances
- Population très dissymétrique : plan de sondage différencié selon la taille des entreprises
- Existence de données fiscales, moins complètes que les données d'enquêtes
- L'exposé s'intéressera au cas des enquêtes annuelles d'entreprise

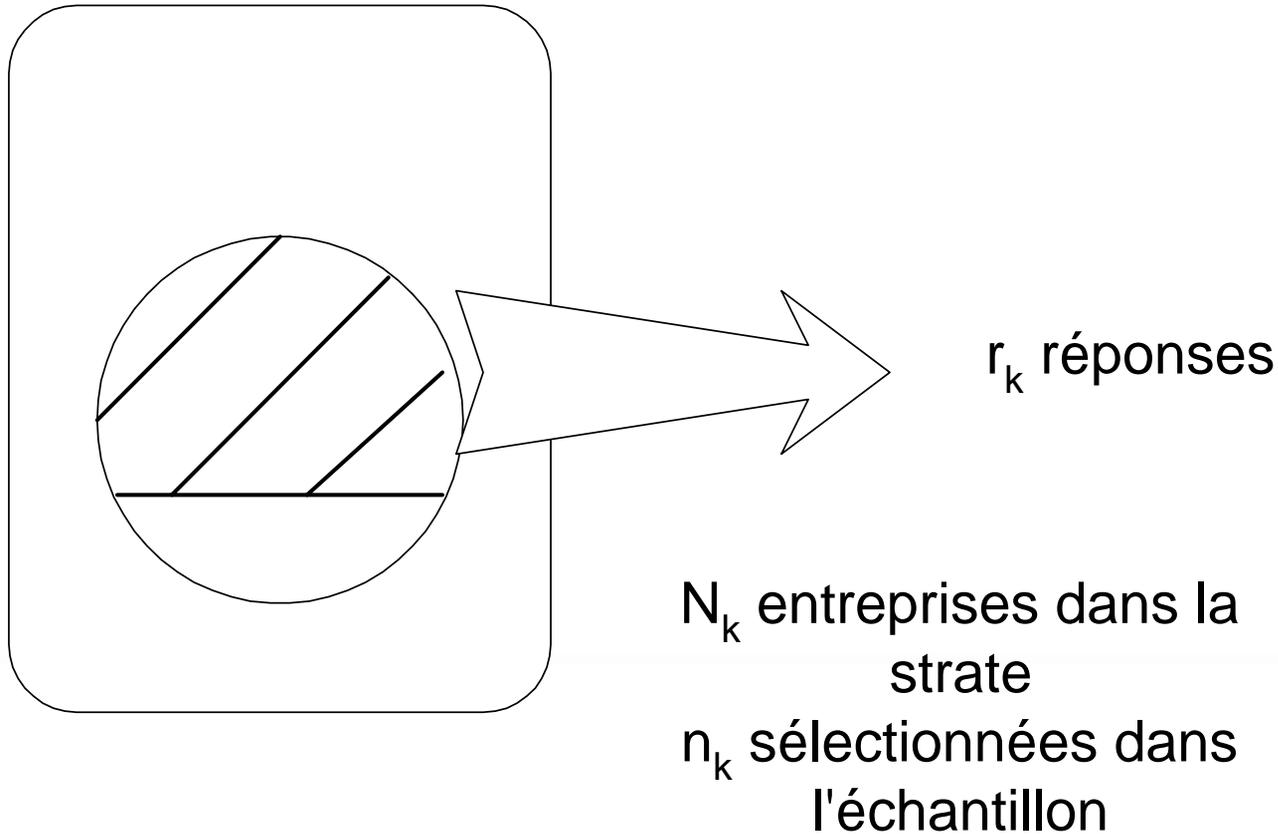


La prise en compte des non-réponses

- Le mode de relance est différencié selon la taille
- Le traitement des non-réponses est lui aussi différencié selon la taille
- Quel est l'impact d'une non-réponse dans chacune des deux catégories ?
- Modélisation du problème pour ...
 - ✓ chiffrer l'importance respective de l'erreur induite par chacune des deux catégories (petites / grandes entreprises)
 - ✓ Fournir des éléments pour définir des priorités dans la relance des non-répondants



Formalisation du problème (1) : estimateur utilisé pour les petites entreprises

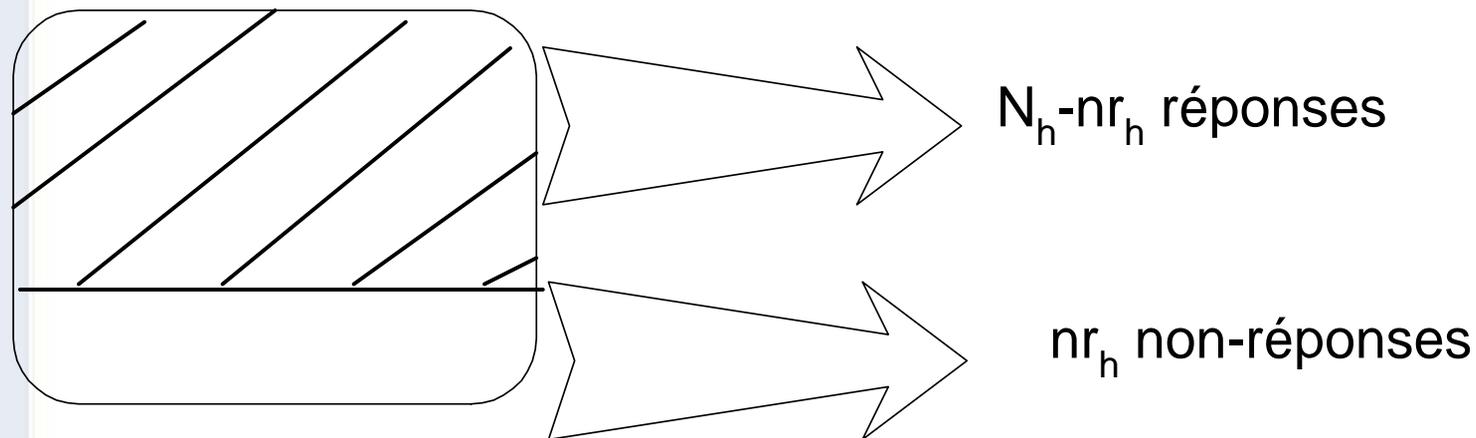


Formalisation du problème (2) : estimateur utilisé pour les petites entreprises

- Estimateur du total d'une variable comme le chiffre d'affaires sur un domaine constitué par un secteur d'activité :

$$\frac{N_k}{r_k} \sum_{rk} Y_i * 1_{APEenq}(i)$$

Formalisation du problème (3) : estimateur utilisé pour les grandes entreprises



Formalisation du problème (4) : estimateur utilisé pour les grandes entreprises

- Pour les entreprises non répondantes, on récupère la valeur du CA dans la source fiscale et on utilise par défaut l'information sur le code APE (activité principale exercée) venant du répertoire : l'estimateur utilisé pour le total d'une variable comme le chiffre d'affaires sur la strate vaut :

$$\sum_{Nh-nrh} Y_i * 1_{APEenq}(i) + \sum_{nrh} Y_i * 1_{APERép}(i)$$



Formalisation du problème (5) : estimateur du total (obtenu sur les deux catégories: petites et grandes entreprises)

- Le mauvais classement éventuel des grandes entreprises non répondantes conduit à un biais
- Les deux catégories (petites et grandes entreprises) apportent de la variance

Formalisation du problème (6) : erreur quadratique moyenne totale

$$\left(\sum_h nr_h A_h\right)^2 + \sum_h \left(nr_h - \frac{nr_h^2}{N_h}\right) B_h^2 + \sum_k \frac{N_k^2}{r_k} \left(1 - \frac{r_k}{N_k}\right) S_k^2$$

➤ Avec :

$$A_h = \frac{1}{N_h} \sum_{N_h} Y_i^* (1_{APErep} - 1_{APEenq})(i)$$

➤ Et : B_h^2 est la variance corrigée de

$$Y_i^* (1_{APErep} - 1_{APEenq})(i)$$



Utilisation des indicateurs calculés

- Utiliser les données de l'année précédente pour décider de publier des résultats, en se positionnant par rapport à un niveau d'erreur anticipé

- Stratégie de relance : introduire des éléments de coût différenciés :
 - ✓ C1 = relance petite entreprise
 - ✓ C2 = visite d'un enquêteur

- Minimiser l'erreur quadratique moyenne sous contrainte de coût



Résultats obtenus sur le secteur du commerce de gros, variable = CA

Classe	Précision obtenue	Taux non-réponses grandes entr.	Taux « stabilité » grandes entreprises
Intermédiaires	0.9%	13%	96.6%
Pr.agricoles bruts	11%	8%	99.3%
Pr. alimentaires	1.2%	11%	98.5%
Biens conso. non alim.	2.4%	14%	97.2%
Pr. interm. non agricoles	1.4%	12%	97.5%
Equip.industr.	1.8%	10%	97.9%
Autres	7.8%	16%	98.3%

Page 11



JMS

18 Mars 2005



Résultats obtenus sur le secteur du commerce de gros, variable = CA

Classe	Part du premier terme (biais au carré)	Part du 2nd terme (variance grandes entr.)	Variance petites entreprises
Intermédiaires	3%	2.7%	94.3%
Pr.agricoles bruts	-	-	100%
Pr. alimentaires	1.2%	4.3%	94.5%
Biens conso. non alim.	1.4%	1.3%	97.3%
Pr. interm. non agricoles	6%	0.5%	93.5%
Equip.industr.	-	0.1%	99.9%
Autres ¹²	0.6%	7.1%	92.3%



Conclusion

- Retour sur les hypothèses
- Affiner la démarche en intégrant des taux de réussite dans les relances
- Développements ultérieurs sur d'autres variables
- La démarche présentée se limite au traitement des non-réponses totales et n'intègre pas d'autres éléments du processus de production des résultats