

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

*Philippe BRION*

*INSEE, Direction des statistiques d'entreprises*

## 1. Introduction

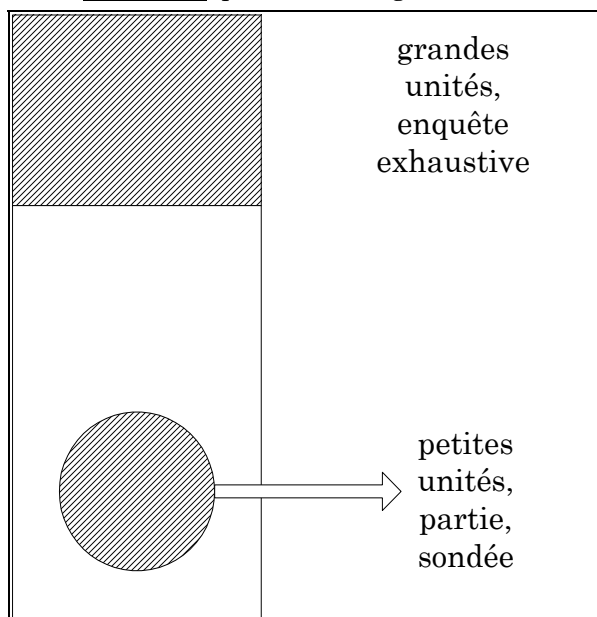
La réalisation des enquêtes menées par le système statistique public auprès des entreprises est un processus complexe, marqué par différentes spécificités. En particulier, ces enquêtes sont en général réalisées par voie postale, ce qui conduit à un échelonnement des retours sur une période assez longue. L'objet de ce papier, centré sur l'enquête la plus importante du dispositif français de statistique d'entreprises, à savoir l'enquête annuelle d'entreprise (EAE [8]), est l'étude de différentes phases du traitement des non-réponses totales, et de leur impact sur la précision des estimations.

La relance des non-réponses totales est différenciée selon la taille des unités non-répondantes : pour les petites unités, des lettres de rappel sont envoyées, alors que pour les grandes, dont les résultats pèsent parfois lourd dans des agrégats sectoriels, un enquêteur peut être envoyé rendre visite à l'entreprise, ce qui est bien plus coûteux.

Le plan de sondage des EAE comporte une partie exhaustive (grandes unités), et une partie sondée (figure 1). De manière simplifiée, la stratification est définie par un croisement entre le secteur d'activité et la taille (en nombre de salariés).

Une caractéristique importante de ces enquêtes tient au caractère comptable des variables demandées dans les questionnaires : ces variables sont également disponibles dans les fichiers des déclarations annuelles de résultat des entreprises (bénéfices industriels et commerciaux, bénéfices non commerciaux), fichiers de la direction générale des impôts auxquels l'Insee a accès. Ceci est dû en particulier à l'existence d'un plan comptable général destiné à différents usages (dont statistique et fiscal). L'amélioration concernant la date de disponibilité de ces fichiers a conduit, ces dernières années, à les utiliser pour traiter le cas des grandes entreprises non-répondantes ; cependant, les fichiers fiscaux ne contiennent pas l'ensemble des variables du questionnaire statistique, et ne peuvent remplacer à l'identique celui-ci.

**Figure 1 :** plan de sondage de l'EAE



On peut donner les grandes lignes du traitement qui est opéré pour les unités non-répondantes (pour une description plus précise, voir [5] ou [8]).

Comme indiqué précédemment, pour les grandes entreprises non-répondantes on cherche à récupérer l'information par la visite d'un enquêteur, ou à défaut par l'utilisation de données fiscales à la place des données d'enquête : cependant, dans ce cas, les données fiscales ne sont pas aussi complètes que celles du questionnaire. En particulier le classement dans un secteur d'activité donné<sup>1</sup> (code APE - activité principale exercée - obtenu à partir de la ventilation du chiffre d'affaires selon les activités élémentaires que l'entreprise indique sur le questionnaire) manque, et doit être approximé par le code APE dont on dispose dans la base de sondage obtenue à partir du répertoire d'entreprises SIRENE. La valeur du code APE a pu évoluer entre-temps. L'impact d'un « mauvais classement » d'une grande entreprise pourra alors être fort sur l'agrégat estimé.

Pour les petites entreprises, le traitement de la non-réponse totale, après la phase de relances postales, est opéré par les méthodes classiques : on procède par imputation dans le cas où on dispose des données de l'année précédente (« reconstitution » du questionnaire à partir de l'application d'un taux d'évolution établi sur les répondantes), ou sinon on utilise la méthode du hotdeck sélectif (à l'intérieur de catégories définies). Pour simplifier les calculs ici, on utilisera dans les sections suivantes un cadre théorique où l'on procède à de la repondération pour traiter les petites entreprises non-répondantes<sup>2</sup>, afin d'évaluer l'impact sur la variance de cette phase de traitement des non-réponses.

*In fine*, la précision obtenue pour un agrégat dépend donc du plan de sondage mis en place d'une part, puis de la prise en compte des non-réponses. La question de l'efficacité du processus de contrôle - redressement des données qui est appliqué est également importante, mais elle ne sera pas abordée ici ; elle a fait l'objet de nombreux articles (voir par exemple [3]), en particulier sur la manière de définir des priorités et des seuils d'arrêt pour ce processus (voir par exemple [4]). Pour la description des méthodes utilisées pour les EAE, on pourra se reporter à [5] ou [8]. On fera donc ici l'hypothèse que le traitement « unitaire » des questionnaires conduit à des données parfaites, où les erreurs d'observation n'existent plus (y compris les non-réponses partielles, qui sont donc censées avoir été correctement imputées).

<sup>1</sup> Qui fait référence à la nomenclature d'activités française (NAF).

<sup>2</sup> La variance estimée par cette méthode a tendance à sous-estimer légèrement la variance obtenue dans le cas où on utilise la méthode du hot-deck, mais à sur-estimer la variance obtenue dans le cadre d'une imputation à partir des données de l'année précédente et d'un taux d'évolution.

Le papier sera donc centré sur la prise en compte des non-réponses totales : quelles actions, en particulier de relance, le statisticien doit-il mettre en place ? Est-il possible de définir des priorités dans les relances, où faut-il faire porter les efforts ?

Pour ce faire, il est nécessaire de quantifier l'impact de différentes phases (échantillonnage, relances opérées de manière différenciée) sur la précision de statistiques. **Le papier prend comme point de départ l'estimation d'agrégats (par exemple le chiffre d'affaires) pour différents domaines de diffusion** (par exemple des secteurs d'activité définis comme les groupes de la NAF, soit le niveau 3 chiffres de cette nomenclature<sup>3</sup>).

La partie suivante de ce papier est consacrée à une modélisation simplifiée<sup>4</sup> du processus global, destinée à évaluer la part d'erreur quadratique moyenne due à chaque composante. Ensuite, est posée la question de l'optimisation du dispositif de relance, partagé entre des relances postales (pour les petites unités, relances peu coûteuses) et des visites sur le terrain (pour les grandes entreprises). Cette question peut être étudiée du point de vue budgétaire, mais également être vue dans la perspective d'une réduction des délais, qui sont en grande partie liés, pour une enquête postale, à la rapidité de récupération des informations. Enfin, une dernière partie présente quelques résultats quantifiés.

## 2. Formalisation du problème

On va s'intéresser à l'estimation d'un agrégat (par exemple le chiffre d'affaires) sur un secteur d'activité donné. Cette estimation va être obtenue comme le total, estimé sur l'ensemble des strates, de la variable chiffre d'affaires multipliée par l'indicatrice d'appartenance, à l'enquête, au secteur en question :  $1_{APE}$ , qui vaut 1 si l'entreprise appartient au secteur, zéro sinon. En effet, on ne peut se contenter de sommer sur les strates définies au départ comme « constituant » ce secteur : pour un certain nombre d'entreprises, l'activité principale a pu évoluer depuis la date d'enregistrement dans le répertoire, et c'est justement un des apports de l'enquête que de « classer » les entreprises de manière correcte.

Dans la suite du papier, on utilisera l'indice  $h$  pour les strates de grandes entreprises, et l'indice  $k$  pour celles des petites entreprises.

### 2.1. Estimateur utilisé pour les grandes entreprises

Pour chaque strate  $h$ , on envoie des questionnaires aux  $N_h$  entreprises de la strate.

On obtient  $N_h - nr_h$  réponses, si le nombre de « grandes non -répondantes » est  $nr_h$ .

On utilise comme contribution de la strate à l'estimateur du total d'une variable  $Y$  (par exemple le chiffre d'affaires sur un secteur) :

$$\sum_{N_h - nr_h} Y_i * 1_{APE_{enq}}(i) + \sum_{nr_h} Y_i * 1_{APE_{rep}}(i)$$

où  $1_{APE_{enq}}$  est l'indicatrice d'appartenance à un secteur d'activité donné d'après le résultat de l'enquête,

et  $1_{APE_{rep}}$  est l'indicatrice d'appartenance au secteur d'après l'information disponible dans la base de sondage (information venant du répertoire d'entreprises).

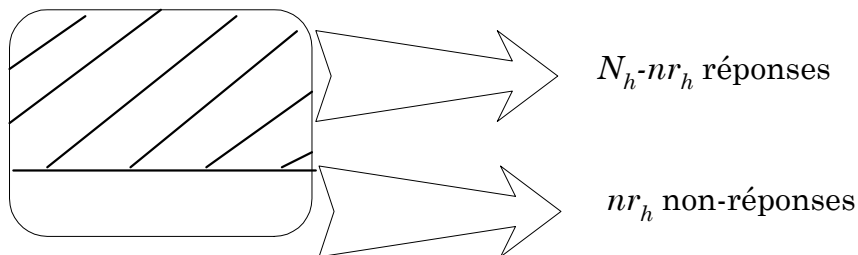
---

<sup>3</sup> Remarquons que le secteur d'activité est utilisé à la fois comme variable stratifiante et pour définir le domaine sur lequel on produit des statistiques ; on pourrait s'intéresser à des domaines de diffusion définis par d'autres critères (par exemple, le chiffre d'affaires des entreprises de plus de cinq ans).

<sup>4</sup> Au sens où la réalité du traitement de l'enquête est nécessairement plus complexe, ce qui demanderait des développements théoriques plus lourds pour qu'elle soit prise en compte « parfaitement ».

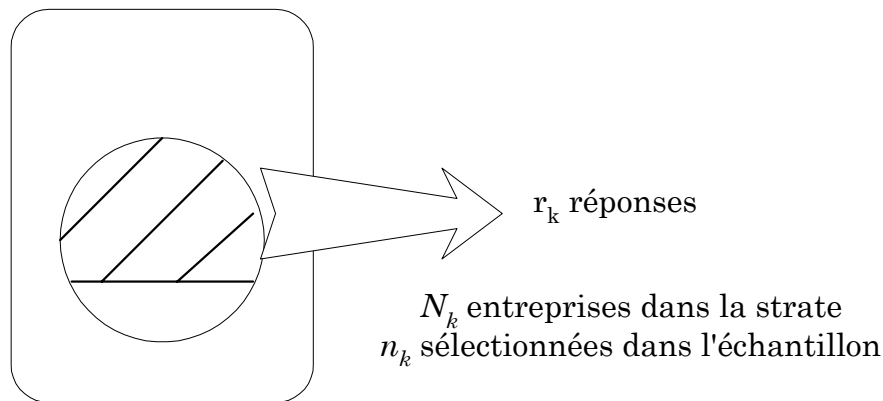
Ce qui signifie que, pour les unités non-répondantes, on s'appuie sur la donnée du répertoire - à défaut d'avoir réalisé l'enquête - pour le classement sectoriel, et qu'on récupère le chiffre d'affaires dans la source fiscale.

**Figure 2 :** strate de grandes entreprises (enquêtée exhaustivement)



## 2.2. Estimateur utilisé pour les petites entreprises

**Figure 3 :** strate échantillonnée



Dans chaque strate  $k$  on a sélectionné  $n_k$  entreprises dans l'échantillon, et obtenu  $r_k$  réponses.

Sous l'hypothèse (volontairement simplificatrice) d'un comportement de non-réponse homogène par strate, on peut utiliser pour chaque strate un estimateur repondéré pour la contribution de la strate à l'estimation du total d'une variable  $Y$  :

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

## 2.3. Biais et variance de l'estimateur d'un total

L'estimateur du total du chiffre d'affaires sur un secteur donné est obtenu à partir de la somme sur l'ensemble des strates de petites et de grandes entreprises :

$$\sum_h \left( \sum_{N_h - nr_h} Y_i * 1_{APEenq}(i) + \sum_{nr_h} Y_i * 1_{APEr\acute{e}p}(i) \right) + \sum_k \left( \frac{N_k}{r_k} \sum_{rk} Y_i * 1_{APEenq}(i) \right)$$

Pour évaluer la précision de cet estimateur, il faut étudier son biais et sa variance.

### 2.3.1. Le biais vient uniquement de l'estimateur relatif aux grandes entreprises

On suppose que le choix des  $nr_h$  unités non-répondantes, parmi les grandes entreprises, résulte d'un mécanisme de tirage au hasard dans la strate  $h$ .

La part d'erreur, due à la strate, pour l'estimation du total de  $Y$ , vaut :

$$\sum_{nr_h} Y_i * 1_{APErep}(i) - \sum_{nr_h} Y_i * 1_{APEenq}(i)$$

L'espérance de cette erreur, soit le biais, vaut :

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

et elle peut être estimée sur l'échantillon par :

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

Au total **sur l'ensemble des strates**, le biais est donc de la forme  $\sum_h nr_h * A_h$ , fonction croissante du nombre de non-répondantes parmi les grandes unités.

### 2.3.2. Les deux catégories (petites et grandes entreprises) contribuent à la variance

#### 2.3.2.1. Petites entreprises

Pour **chaque strate**  $k$ , la variance de l'estimateur du total vaut :  $\frac{N_k^2}{r_k} (1 - \frac{r_k}{N_k}) S_k^2$ , sous l'hypothèse, toujours, d'un comportement de non-réponse homogène à l'intérieur de la strate, où  $S_k^2$  est la variance corrigée de la variable  $Y_i * 1_{APEenq}(i)$  sur la strate  $k$ .

#### 2.3.2.2. Grandes entreprises

L'estimateur du total sur la strate  $h$  peut être écrit comme :

$$\sum_{N_h} Y_i * 1_{APEenq}(i) + \sum_{nr_h} Y_i * (1_{APErep} - 1_{APEenq})(i)$$

Le premier terme étant une constante, la variance de cet estimateur du total **sur la strate**  $h$  vaut :

$$V(\sum_{nr_h} Y_i * (1_{APErep} - 1_{APEenq})(i)) = nr_h (1 - \frac{nr_h}{N_h})^2 B_h^2$$

où :

$B_h^2$  est la variance corrigée de la variable  $Y_i * (1_{APErep} - 1_{APEenq})(i)$

et peut être estimée sur l'échantillon des  $N_h - nr_h$  réponses.

## 2.4. Erreur quadratique moyenne totale

Si l'on somme le carré du biais et la variance obtenue sur chacune des strates (de petites ou de grandes entreprises), on obtient la valeur suivante pour l'erreur quadratique moyenne de l'estimateur du total de la variable sur un secteur donné :

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

les deux premiers termes étant sommés sur les  $H$  strates de grandes entreprises, et le troisième sur les  $K$  strates de petites entreprises.

On remarque que le premier terme est lié au carré du nombre de non-répondantes « grandes entreprises ».

## 3. Conséquences sur le processus de relances

### 3.1. Minimisation de l'erreur quadratique moyenne sous contrainte

On peut se poser la question des priorités à définir pour les actions de relance à mener auprès des entreprises non-répondantes. Cette question a été traitée dans certains papiers (voir par exemple [6] ou [7]), mais plus sous la forme d'un « score » individuel affecté à chaque entreprise, qui va permettre d'établir un classement des unités non-répondantes. L'idée développée ici est plutôt celle de la mise au point d'indicateurs relatifs aux différentes strates : sur lesquelles faut-il porter l'effort, est-il possible de décider d'arrêter la relance sur d'autres catégories ? Cette problématique a déjà été abordée dans [1], il s'agit ici de la prolonger en tenant compte du traitement différencié des petites et des grandes entreprises.

Au niveau des relances, on peut agir sur deux ensembles de paramètres :  $nr_h$ , le nombre de grandes non-répondantes dans chaque strate (en faisant appel à des visites d'enquêteurs), et  $r_k$ , le nombre d'unités répondantes (sous forme de relances postales ou téléphoniques).

Les relances par enquêteur ont un coût  $C_1$  et les relances postales un coût  $C_2$  (plus faible).

La contrainte de budget est de la forme :

$$\sum_h C_1(nr_{h0} - nr_h) + \sum_k C_2(r_k - r_{k0}) = C$$

où  $nr_{h0}$  est le nombre de grandes non-répondantes obtenu « de manière spontanée »,

et  $r_{k0}$  est le nombre de répondantes, parmi les petites entreprises, obtenu également de manière spontanée.

La quantité à minimiser est l'erreur quadratique moyenne :

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

avec  $0 \leq nr_h \leq nr_{h0}$

et  $r_{k0} \leq r_k \leq n_k$ .

### 3.2. Résolution

Elle se fait en écrivant le Lagrangien :

$$L = \left(\sum_h nr_h A_h\right)^2 + \sum_h \left(nr_h - \frac{nr_h^2}{N_h}\right) B_h^2 + \sum_k \frac{N_k^2}{r_k} \left(1 - \frac{r_k}{N_k}\right) S_k^2 + \lambda \left(\sum_h C_1 nr_h - \sum_k C_2 r_k\right)$$

En dérivant par rapport à  $r_k$  :

$$\frac{S_k^2}{2} N_k^2 + \lambda C_2 = 0$$

En dérivant par rapport à  $nr_h$  :

$$2\left(\sum_h nr_h A_h\right) A_h + B_h^2 \left(1 - 2 \frac{nr_h}{N_h}\right) + \lambda C_1 = 0$$

La résolution est décrite en annexe : la contrainte de budget va conduire à une première répartition entre le budget global de relances par enquêteur et celui lié aux relances postales, et ensuite on obtient le nombre de relances au sein de chaque strate, qui pour les strates de petites unités « retrouve » l'allocation optimale de Neyman.

Comment mettre en pratique la méthode une année donnée ? On peut utiliser les estimations chiffrées des différents éléments  $A_h$ ,  $B_h$ ,  $S_k$  obtenues à partir des données de l'enquête de l'année précédente, et injecter les paramètres correspondant aux nombres de réponses constatés à un moment donné de l'enquête. La plus ou moins grande stabilité des strates de grandes entreprises (en termes de classement sectoriel) et l'homogénéité relative de chaque strate de petites entreprises seront ainsi prises en compte pour fournir des nombres d'unités à relancer dans chacune des strates. Si l'enquête est suffisamment avancée, on peut procéder à des estimations des  $A_h$ ,  $B_h$ ,  $S_k$  à partir des données de l'année en cours.

On peut également se poser la question en termes de délais : faut-il relancer les unités non-répondantes jusqu'à ce qu'on ait récupéré l'ensemble des données ? Le modèle présenté ci-dessus permet d'obtenir une information quantifiée concernant la précision (ici, l'erreur quadratique moyenne) à tout moment de l'enquête : on peut décider de publier des résultats pour un secteur donné si, par exemple, l'erreur quadratique moyenne reste dans une limite raisonnable par rapport à celle qu'on aurait eue si on avait obtenu l'ensemble des réponses (on peut par exemple fixer une marge de 10% supplémentaires). Pour produire une estimation à ce niveau de diffusion, on n'a alors plus besoin de relancer les unités non-répondantes ; par contre, si l'on s'intéresse à des niveaux de diffusion plus fins, il faudra sans doute relancer de manière plus complète. Le modèle s'inscrit ainsi dans la problématique de définition d'un critère d'arrêt d'enquête.

## 4. Résultats quantifiés

Des calculs ont été menés<sup>5</sup> sur le secteur du commerce de gros, pour l'EAE menée sur l'exercice 2003. Pour cette enquête, le seuil d'exhaustivité est en général situé à 20 salariés, des exceptions existant pour certains sous-secteurs ; les taux de sondage sont compris entre 1/30 et 1/2 pour les strates sondées.

Le tableau 1 présente quelques résultats relatifs aux groupes (niveau 3 chiffres de la nomenclature NAF) : précision calculée comme la racine carrée de l'erreur quadratique moyenne finale de l'estimateur rapportée à la grandeur estimée (en l'occurrence le chiffre d'affaires du groupe considéré), taux de non-réponses final constaté sur les grandes entreprises, et taux de stabilité des grandes entreprises, calculé (à partir des unités répondantes) comme le pourcentage d'unités toujours classées à l'enquête dans le secteur considéré (par rapport à celles qui étaient classées dans ce secteur au lancement de l'enquête).

---

<sup>5</sup> Je remercie Sébastien Hallépée et Thierry Mainaud, de la direction des statistiques d'entreprises de l'Insee, pour les calculs réalisés sur les données de l'EAE.

**Tableau 1** : éléments sur l'EAE commerce de gros (exercice 2003)

Classe	Précision obtenue à l'enquête	Taux de non-réponses grandes entreprises	Taux de stabilité grandes entreprises
Intermédiaires	0.9%	13%	96.6%
Produits agricoles bruts	11%	8%	99.3%
Produits alimentaires	1.2%	11%	98.5%
Biens consommation non alimentaires	2.4%	14%	97.2%
Produits intermédiaires non agricoles	1.4%	12%	97.5%
Equipements industriels	1.8%	10%	97.9%
Autres	7.8%	16%	98.3%

La précision est comprise entre 1% et 2% pour la plupart des secteurs, le taux de réponse des grandes unités se situant globalement entre 10 et 15%.

Le tableau 2 donne la décomposition de l'erreur quadratique moyenne selon les trois termes détaillés dans la partie 2 :  $(\sum_h nr_h A_h)^2$ ,  $\sum_h (nr_h - \frac{nr_h^2}{N_h}) B_h^2$ ,  $\sum_k \frac{N_k^2}{r_k} (1 - \frac{r_k}{N_k}) S_k^2$ .

**Tableau 2** : Décomposition de l'erreur quadratique moyenne selon les trois termes (EAE commerce de gros, exercice 2003)

Classe	1 <sup>er</sup> terme	2 <sup>ème</sup> terme	3 <sup>ème</sup> terme
Intermédiaires	3%	2.7%	94.3%
Produits agricoles bruts	-	-	100%
Produits alimentaires	1.2%	4.3%	94.5%
Biens consommation non alimentaires	1.4%	1.3%	97.3%
Produits intermédiaires non agricoles	6%	0.5%	93.5%
Equipements industriels	-	0.1%	99.9%
Autres	0.6%	7.1%	92.3%

*Note de lecture : la somme en ligne fait 100%.*

La lecture de ce tableau 2 montre que la part de l'erreur quadratique due au mauvais classement de certaines grandes entreprises reste limitée : d'une certaine façon, l'approximation qu'on fait parfois en considérant que, ces strates étant exhaustives, elles n'apportent pas de variance, paraît à peu près justifiée. Cependant, si le taux de non réponses des grandes unités devait évoluer à la hausse, la part due à ces strates pourrait devenir nettement plus forte (en particulier le 1<sup>er</sup> terme est lié au carré du nombre de non-répondantes).

On voit également que certains secteurs sont sans doute plus « instables » que d'autres, et que c'est sur ceux-là qu'il faut porter les efforts de relance afin de ne pas trop dégrader les résultats.



## 5. Conclusion

La modélisation présentée ici, même si elle repose sur plusieurs hypothèses fortes, peut permettre d'avancer dans la mise en place d'une démarche globale de traitement de l'enquête. Elle peut être affinée en anticipant un taux de réussite dans les différents types de relance (ceci peut être introduit au niveau des coûts unitaires). On pourrait également prolonger la démarche en étudiant certaines variables, autres que celles du compte de résultat (ces dernières étant disponibles à la fois dans l'enquête et dans la source fiscale), qu'on est conduit à « reconstruire », pour les grandes non-répondantes, quand on utilise la source fiscale, par exemple en appliquant un ratio moyen (calculé sur les unités répondantes de la même catégorie) à une variable fiscale.

Par ailleurs, il est important de revenir sur les hypothèses faites lors des calculs : elles ont consisté à supposer des comportements de non-réponses homogènes dans les strates. Il est essentiel de contrôler cette hypothèse, à partir des données dont on dispose dans la base de sondage, avant de publier des estimations précoces, par exemple.

La démarche permet d'étayer les choix relatifs à la gestion temporelle du processus de relance, en indiquant les strates où l'effort de relance doit être produit, mais elle ne prétend pas régler tous les problèmes liés à la production des résultats : d'autres éléments sont à prendre en compte, comme par exemple la validation de ces résultats à l'aide d'autres sources, ou encore les problèmes posés par les unités statistiques complexes (groupes avec l'existence de flux internes importants, par exemple), pour lesquelles on procède à des traitements spécifiques [2].

## Bibliographie

- [1] Brion Ph., "Arbitrages relatifs aux contraintes de réalisation d'une enquête. Faut-il s'acharner ? Etude de deux exemples", in *Echantillonnage et méthodes d'enquêtes*, Dunod, 2004
- [2] Dervieux L., "The statistical units by example", 17ème table ronde internationale sur les répertoires d'entreprises, Rome (disponible sur <http://petra1.istat.it/17roundtable>), 2003
- [3] Granquist L., Kovar J., "Editing of survey data : how much is enough ?", in *Survey measurement and process quality*, Lyberg et al. ed., J. Wiley and sons, New York, 1997
- [4] Lawrence D., McKenzie R., "The general application of significance editing", *Journal of Official Statistics*, vol. 16, n°3, pp. 243-253, 2000
- [5] Mauguin J., "Les procédures automatiques de contrôle de données dans les enquêtes annuelles d'entreprise", in *Echantillonnage et méthodes d'enquêtes*, Dunod, 2004
- [6] McKenzie R., "A framework for priority contact of non respondents", *Second International Conference on Establishment Surveys*, Buffalo, 2000
- [7] Poirier C., Philips R., Pursey S., "The use of a score function in a data collection context", *Working paper n°28, Conference of European Statisticians, UN/ECE Work Session on Statistical Data Editing*, Madrid, 2003
- [8] Rivière P., "Enquêtes annuelles d'entreprise : à la rencontre du 4e type", *Courrier des statistiques*, n° 78, août 1996

## Annexe : Minimisation de la variance d'estimation sous contrainte (partie 3.2)

Sous réserve que les contraintes ne soient pas saturées, on détermine les quantités  $nr_h$  en fonction de  $\lambda$ , en partant de la deuxième série d'équations établie au 3.2. :

$$2\left(\sum_h nr_h A_h\right)A_h + B_h^2\left(1 - 2\frac{nr_h}{N_h}\right) + \lambda C_1 = 0$$

Si l'on multiplie chaque membre de gauche par  $\frac{1}{2} \frac{A_h N_h}{B_h^2}$ , et si on somme sur l'ensemble des strates  $h$ , on trouve :

$$\left(\sum_h nr_h A_h\right)\left(\sum_h \frac{N_h A_h^2}{B_h^2} - 1\right) = -\frac{\lambda C_1}{2} \sum_h \frac{A_h N_h}{B_h^2} - \sum_h \frac{A_h N_h}{2}$$

Comme par ailleurs  $\frac{S_k^2}{2} N_k^2 + \lambda C_2 = 0$ , on peut en déduire  $\frac{\sum_k S_k N_k}{\sum_k r_k} = \sqrt{-\lambda C_2}$

En égalisant sur  $\lambda$ , on peut exprimer  $\left(\sum_h nr_h A_h\right)$  en fonction de  $\sum_k r_k$ .

Grâce à l'équation du haut de la page, on exprime ensuite  $nr_h$  en fonction de  $\left(\sum_h nr_h A_h\right)$ , et en revenant à la fonction de coût initiale on en déduit la valeur chiffrée des différents paramètres.