

L'UTILISATION DU CALAGE SUR MARGES POUR LES ENQUÊTES THÉMATIQUES DU SESSI : UN RETOUR D'EXPÉRIENCE

Boris GUANNEL, Yann LHOMME

Sessi, sous-direction Etudes

Le Sessi est confronté à partir de la fin des années 90 à une dégradation soudaine de la qualité des enquêtes thématiques. Ces enquêtes légères sont destinées à apporter des éclairages particuliers, par exemple sur le financement de l'innovation, les échanges intra-groupes.... Elles sont pénalisées par une précision moindre des estimations, une non-réponse importante et une désaffection croissante des grandes entreprises.

La post-stratification alors traditionnellement utilisée comme méthode de redressement assurait une représentativité des échantillons en nombre d'entreprises¹. Mais les résultats d'enquête restituaient imparfaitement le poids économique de chaque secteur de l'industrie. La post-stratification est adaptée aux enquêtes qualitatives à bon taux de réponse ; elle est inadaptée pour des enquêtes quantitatives à faible taux de réponse.

Le Sessi a engagé une réflexion sur les méthodes de redressement d'enquête. Il a systématisé l'utilisation d'informations auxiliaires grâce à la procédure de calage sur marges. Cette méthode, couramment utilisée à l'Insee sur les enquêtes ménages, fait appel au logiciel Calmar (calage sur marges), développé à cet usage. L'échantillon est redressé en utilisant des informations auxiliaires, appelées variables de calage, qui peuvent être aussi bien des dénombrements au sein de variables catégorielles que des variables quantitatives. Cette méthode présente de nombreux avantages (partie 1).

La non-réponse introduit un biais dans les estimations. Le redressement est donc précédé d'une analyse des entreprises non-répondantes. Cette non-réponse est éventuellement corrigée avant le calage. Ce dernier peut ainsi être réalisé dans de bonnes conditions (partie 2).

Pour être de bonne qualité, le redressement respecte certains critères : choix adapté des variables de calage et examen de la déformation finale de l'échantillon (partie 3). Un exemple de redressement est donné en partie 4.

A ce jour, le Sessi bénéficie de l'expérience du redressement par calage de quatre enquêtes thématiques successives (*annexe I*). Les enquêtes sur le financement de l'innovation (Fit, 1999) et sur les services aux entreprises (1999) constituent une première utilisation des procédures de redressement. L'enquête sur les échanges internationaux intra groupe (EIIG, 2000) se caractérise par une non-réponse particulière qui a entraîné des traitements spécifiques. L'enquête communautaire sur l'innovation (CIS3, 2001) intègre l'expérience de ces trois premières enquêtes et introduit de nouvelles méthodes de gestion de la collecte.

¹ La post-stratification est une méthode de redressement par re-pondération. Elle consiste à donner à chaque entreprise un poids qui est le produit de l'inverse du taux de sondage et du taux de réponse, pour les strates croisant tranche d'effectif et secteur.

1. De l'intérêt du calage sur marges.

1.1 Augmenter la précision des estimateurs

Le Sessi dispose d'informations précises sur les entreprises enquêtées. Cette information auxiliaire provient par exemple de l'enquête annuelle d'entreprise (EAE). Pour les enquêtes thématiques, elle est utilisée au moment du tirage de l'échantillon pour définir les strates.

Lors du redressement, l'information auxiliaire sert à augmenter la précision des estimateurs. Si elle est bien corrélée avec la variable d'intérêt, l'intervalle de confiance est ainsi diminué. De plus, l'estimateur est le moins biaisé possible.

Le calage est une méthode de redressement particulière. Pour cela, elle utilise des informations auxiliaires corrélées à la variable d'intérêt pour augmenter la précision des estimateurs. Le calage consiste à faire coïncider la structure de l'échantillon avec celle de la population.

1.2 La possibilité de réduire conjointement le biais de non-réponse

Sous certaines conditions le redressement par calage permet conjointement de corriger la non-réponse.

La non-réponse nuit à la qualité des enquêtes car elle introduit des biais. Ainsi, dans l'enquête EIIG, les entreprises qui échangent peu de produits manufacturés sont bien moins répondantes que celles fortement présentes à l'international. Ces deux types d'entreprises n'ont pas le même comportement en terme d'ouverture commerciale à l'extérieur. Ne pas corriger la non-réponse revient à déformer l'échantillon : l'impact des réponses des entreprises fortement internationalisées par rapport à celles qui le sont moins est sur-estimé.

Le calage permet de construire des estimations corrigées du biais de non-réponse. En modifiant l'échantillon pour le contraindre à adopter la structure de la population globale, il tient compte des spécificités de cette population et assure de fait une « représentativité » des résultats.

Sous certaines conditions, cette meilleure « représentativité » traduit en fait une réduction du biais de non-réponse. Cet avantage n'est effectif que si certaines conditions sont respectées : les variables de calage doivent être corrélées avec celles qui déterminent la non-réponse.

Dans ce cas, le calage sur marge permet théoriquement de réduire le biais de non-réponse. C'est un net avantage. Certains retraitements de la non-réponse globale peuvent en effet s'avérer fort coûteux, tel celui de l'enquête sur les échanges intra-groupes (partie 2).

1.3 Améliorer le cadrage macroéconomique des enquêtes, sur plusieurs agrégats et en même temps

Les résultats d'enquêtes thématiques font référence à des grandeurs macroéconomiques connues. Par exemple, l'enquête sur le financement de l'innovation sert entre autres à analyser les politiques publiques d'aides à l'innovation. Obtenus par l'enquête à un niveau global, les montants des aides, par exemple de l'Agence nationale de valorisation de la recherche (Anvar), doivent être assez peu éloignés de ceux connus par ailleurs.

Les enquêtes sont donc calées selon des grandeurs macroéconomiques utiles à l'interprétation, de manière à ce que les résultats cadrent avec la réalité économique.

Le calage sur marges permet de redresser l'enquête sur plusieurs variables en même temps. Au contraire, la post-stratification ne fait coïncider l'échantillon qu'avec une seule variable.

1.4 Une indication du niveau d'interprétation des résultats

Le calage sur marges indique quelles sont les limites d'interprétation des résultats. La convergence n'est pas assurée pour certains niveaux trop détaillés. L'enquête sur les Services n'est ainsi « représentative » qu'au niveau NES36 ou en quatre tranches de taille, ces deux dimensions ne pouvant être croisées.

Cette information quant aux limites des interprétations accompagne la diffusion des fichiers auprès des utilisateurs extérieurs.

Contrairement à l'imputation², le calage sur marges permet d'indiquer le niveau d'interprétation des estimations redressées, sans craindre de les voir biaisées par la non-réponse. Ainsi, dans le cas de l'enquête EIIG, théoriquement exhaustive, imputer les réponses d'entreprises manquantes permettrait de faire des estimations directement à partir des réponses des entreprises, sur de petits nombres d'entreprises, voire au niveau individuel.

Compte tenu du faible taux de réponse de cette enquête, cette utilisation sur de petits ensembles conduirait à des estimations biaisées, très peu précises. Cette limitation peut échapper à l'utilisateur qui est bien souvent tenté d'utiliser l'enquête à un niveau de détail très fin. Il oublie ainsi que la probabilité est forte d'avoir sur de petits échantillons une entreprise dont les résultats sont imputés et qui influence fortement les estimations.

1.5. Un instrument de réflexion

1.5.1 Agir sur la non-réponse lors de la phase de collecte

Mise en place avec l'enquête communautaire sur l'innovation (2001), l'analyse de la non-réponse est effectuée le plus en amont possible lors de la phase de collecte. Elle conditionne en effet la qualité du calage final. Elle est donc optimisée pour que l'étape du redressement se déroule dans les meilleures conditions possibles.

Pour cela, en cours d'enquête, un premier calage est effectué. Il se fait dans deux dimensions : selon le chiffre d'affaires et selon un dénombrement d'entreprise, en fonction des strates qui croisent taille et secteur. Cela permet de repérer les populations d'entreprises qui globalement présentent des taux de réponse ou des taux de couverture de chiffre d'affaires plus faibles et qui de ce fait déforment fortement l'échantillon lors du redressement.

Ce premier calage est présenté dans un tableau de bord, qui chaque semaine fournit aussi les taux de réponse et les taux de couverture de chiffre d'affaires par strate. Ainsi, le tableau de bord permet de localiser les strates à faible taux de réponse ou à bas taux de couverture en chiffre d'affaires. Le calage intègre conjointement ces deux dimensions : il localise donc ces mêmes strates mais permet aussi de repérer les strates où les taux de réponse et les taux de couverture en chiffres d'affaires sont moyens.

Lors du pilotage des enquêtes thématiques, l'un des objectifs est de faire des relances ciblées. Sont donc relancées en priorité les entreprises des strates affectées de faible taux de réponse ou de faible taux de couverture. Les entreprises de ces strates seraient fortement déformées par le calage.

Ainsi, la non-réponse est homogénéisée : le taux de réponse est voisin pour les strates qui servent au calage final.

Cette approche globale se double d'une approche individuelle. Lors de ce premier calage, les grandes entreprises qui ont des poids importants font l'objet d'une analyse spécifique. Il faut en effet éviter que certaines entreprises, Renault par exemple, soient affectées de poids importants, parce que les petites et moyennes entreprises du secteur automobile sont peu répondantes. Ces strates font alors aussi l'objet de relances spécifiques.

² Elle consiste à attribuer une réponse estimée aux entreprises non répondantes.

1.5.2 Réagir rapidement aux résultats d'enquête

Caler rapidement en cours d'enquête a un autre intérêt : il permet de suivre le plus en amont les résultats d'enquête, d'un point de vue macroéconomique.

Généralement, les estimations redressées en cours de collecte convergent rapidement vers leurs valeurs définitives. Il est ainsi possible de disposer rapidement de résultats provisoires robustes, qui permettent de faire face aux demandes d'informations en cours d'enquête.

Ces demandes peuvent être politiques, elles peuvent aussi intéresser le statisticien. Dans le cadre de l'enquête communautaire sur l'innovation, le taux d'entreprises innovantes constitue une donnée à peu près connue, notamment parce que c'est la 4^{ème} enquête sur ce thème. Constatant une diminution incite à examiner rapidement les différentes hypothèses qui pourraient expliquer ce phénomène, et sur lesquelles on peut encore agir en cours de collecte. Ainsi, un rappel des entreprises a permis d'éliminer l'hypothèse d'une erreur de mesure.

1.5.3 Déceler les problèmes d'échantillonnage

Caler revient à faire converger la structure de la base d'échantillonnage vers celle d'une base de calage. Ces deux bases sont en général différentes. Au Sessi, le tirage de l'échantillon se fait à partir d'un fichier unique de lancement d'enquête, qui est un répertoire d'entreprises. Il est mis à jour en temps réel. Les marges elles sont fabriquées à partir de l'enquête annuelle d'entreprise (EAE), à finalité plus économique. L'EAE est disponible une année après le lancement de l'enquête thématique.

L'une des conséquences de ce décalage temporel conduit à enquêter des entreprises, qui entre-temps disparaissent. Lors du calage, ces firmes sont éliminées.

Des différences de concepts existent également entre les deux bases. Cela a conduit à des défauts de couverture. Par exemple, l'enquête Services, la première à avoir été calée au Sessi, ne convergerait pas. Pour stratifier l'échantillon, la base d'échantillonnage du Sessi utilise l'effectif de fin d'année qui correspond aux unités légales. Quelques grandes entreprises de l'industrie automobile n'ont donc pas été interrogées. En effet, dans le fichier de lancement, elle ne disposent d'aucun effectif, les salariés étant prêtés par une autre entreprise du groupe. Au contraire, l'EAE dispose d'un effectif corrigé, qui correspond à une réalité économique.

Par la suite, les plans de sondage ont intégré ces effectifs corrigés, afin que base d'échantillonnage et base de calage soient identiques.

2. Préalablement à la mise en œuvre du calage, le traitement de la non-réponse

Le calage sur marges n'est pas une technique qui permet automatiquement de corriger les effets de la non-réponse et de redresser une enquête : mal utilisé, il peut détériorer la qualité des estimations. Avant le calage, il faut donc analyser les déterminants de la non-réponse.

2.1 Traitement conjoint de la non-réponse et du redressement

Lors du calage, réaliser en une seule étape le redressement de l'enquête et le traitement de la non-réponse repose sur une hypothèse particulière concernant la non-réponse. Cela revient à considérer que cette dernière est expliquée par les variables de calage.

C'est bien souvent le cas pour les enquêtes thématiques du Sessi, à l'exception de l'enquête sur les échanges internationaux traitée ci-après. La non réponse dépend en général de la taille des entreprises et de leur secteur d'activité, qui sont par ailleurs les variables de calage.

2.2 Le traitement particulier de la non-réponse dans l'enquête sur les échanges intragroupe

L'enquête EIIG est caractérisée par une hétérogénéité de la non-réponse et surtout de ses causes. Une étude approfondie montre des caractéristiques différentes des entreprises « non répondantes car non enquêtées » et des entreprises « enquêtées qui n'ont pas pu ou voulu répondre ». Il faut donc redresser ces deux types de non-réponse.

Une typologie de ces deux populations a pu être dressée :

- Les entreprises « non répondantes car non enquêtées » sont d'une part celles qui n'ont pas pu être repérées au lancement de l'enquête, compte tenu de leur petite taille et donc de leur manque de visibilité (elles ne sont que rarement sujettes aux enquêtes). D'autre part, les entreprises des IAA et du commerce de gros ont été mieux repérées que celles de l'industrie. Enfin, les entreprises appartenant à des groupes étrangers ont été enquêtées plus exhaustivement que les entreprises françaises. Cela s'explique par l'attention particulière dont elles ont fait l'objet lors de la collecte, afin de permettre des exploitations spécifiques par la suite.
- De même, parmi les entreprises enquêtées, celles qui répondent le mieux sont celles qui sont les plus visibles, compte tenu de l'importance de leurs flux douaniers. Les firmes interrogées par courrier sont mieux répondantes que celles enquêtées en face à face : ces dernières étaient les plus problématiques. Enfin, les secteurs industriels sont moins bien répondants que ceux des IAA et du commerce de gros, avec une non-réponse plus prononcée pour les secteurs de l'industrie des biens de consommation (EC) et de l'énergie (EG).

Un traitement de la non-réponse indépendant du calage a été fait, afin de faciliter ce dernier. Il a été réalisé par la méthode des Groupements de Réponses Homogènes (GRH, cf. *annexe 2*). Cette méthode consiste à déterminer, a posteriori, de nouveaux poids d'échantillonnage obtenus en modélisant la probabilité pour une entreprise donnée d'être ou non répondante.

Pour ce faire, on commence par trouver des variables discrètes expliquant qu'une entreprise soit enquêtée ou non, puis la probabilité pour une entreprise enquêtée d'être répondante ou non. Les déterminants de ces deux phénomènes étant distincts, le fait d'être une entreprise répondante a été modélisé comme le produit des deux phénomènes indépendants : être enquêtée ou non et répondre ou non au questionnaire.

En croisant les variables discrètes expliquant la non-réponse, on génère ainsi des groupes homogènes d'entreprises répondantes. La probabilité pour une entreprise d'un de ces groupes d'être finalement répondante est estimée par le produit de la probabilité d'être enquêtée et de la probabilité d'être répondante, pour les entreprises répondantes. Le rapport n/N , avec N la taille réelle du groupe et n le nombre d'entreprises de ce groupe qui sont effectivement enquêtées (resp. répondantes), est une approximation de la probabilité pour une entreprise du groupe d'être enquêtée (resp. répondante).

On définit ainsi un ensemble de coefficients qui sont comparables aux probabilités d'inclusion d'un tirage aléatoire simple. Le biais de non-réponse est ainsi diminué.

3. La mise en œuvre du calage sur marges

Lors des redressements par calage sur marges des enquêtes d'entreprises, l'objectif essentiel est l'attribution de coefficients raisonnables et appropriés à chaque entreprise, afin de ne pas trop déformer l'échantillon et les résultats d'enquête.

3.1 Choix des variables de calage

Les sources dont proviennent les variables de calage doivent être d'une plus grande précision que l'enquête. Elles doivent donc soit être exhaustives, soit provenir d'enquêtes dont le taux d'échantillonnage est significativement plus élevé.

De plus, les informations mobilisées doivent porter sur la même période que l'enquête à redresser. Au Sessi, les bases de calage sont en général constituées d'informations provenant des Enquêtes Annuelles d'Entreprise (EAE) : elles sont exhaustives et portent sur la même période. D'autres sources ont également été mobilisées : par exemple les fichiers des douanes ou les fichiers administratifs de gestion des aides à l'innovation.

Le choix des variables de calage dépend des thèmes d'intérêt de l'enquête. Ainsi, dans le cadre de l'enquête sur les services, le calage assure que les résultats sont cohérents avec les comportements d'achats de services identifiés dans l'EAE.

Le Sessi, qui gère les enquêtes annuelles d'entreprises, cherche aussi à ce que les résultats d'enquête soient représentatifs de l'industrie, que ce soit en nombre d'entreprises ou en chiffre d'affaires.

Le choix des variables dépend des corrélations qui pourraient exister entre elles. Il est nécessaire que les variables de calage ne soient pas corrélées entre elles.

En premier lieu, la procédure de calage sur marges ne fonctionnerait pas convenablement.

En deuxième lieu, choisir une seule variable de calage parmi deux corrélées permet d'atteindre une bonne valeur cible pour la seconde variable. Par exemple, pour l'enquête sur les échanges intragroupe, le calage du montant des importations et des exportations des entreprises s'est fait par secteur et a été préféré à celui par secteur et produit.

Enfin, dans le cas où la non-réponse est corrigée lors du calage, les variables de calage doivent être corrélées à la non-réponse et à la variable d'intérêt.

3.2 La première étape du calage

Le calage se fait par tâtonnements itératifs. Ainsi, pour les quatre enquêtes calées au Sessi, le premier calage vise un faible niveau de détail, une répartition en NAF 16 par exemple. Puis il s'affine peu à peu, une fois toutes les contraintes identifiées et résolues.

Lors de cette première étape, les différences de structure entre la base de calage et la base à caler sont examinées. Cela permet d'évaluer le niveau de déformation auquel sera soumis l'échantillon lors de la convergence vers les marges. Ces différences sont dues à la non-réponse, lorsque cette dernière est corrigée dans cette même étape. Mais elles proviennent aussi de défauts d'échantillonnage.

Si les contraintes sont trop fortes, le calage convergera difficilement. L'alternative consiste alors à caler plus grossièrement l'enquête, à un niveau moins fin que celui initialement envisagé.

3.3 Amélioration du calage

Une fois ce premier calage effectué, il est possible de l'améliorer. Pour cela, il existe différents critères de qualité. Le principal consiste à étudier individuellement les déformations apportées par le calage : les poids finaux après calage sont comparés aux poids initiaux, corrigés ou non de la non-réponse.

3.3.1 Faible amplitude des rapports de poids

La distribution des rapports entre poids finaux et poids initiaux de chaque entreprise doit être de faible amplitude et concentrée autour de 1³.

Ce critère permet de s'assurer que la déformation de l'échantillon calé par rapport à l'échantillon enquêté initial est raisonnable. En effet, les estimateurs calculés à partir des pondérations initiales d'échantillonnage sont en général sans biais.

Le calage sur marge, en modifiant ces pondérations, modifie en même temps le caractère sans biais des estimations. Aussi, pour que ce biais soit négligeable, il faut que les rapports de pondérations soient proches de un, c'est à dire que les pondérations après calage ne sont pas très éloignées des pondérations initiales.

Il faut, en outre que la distribution des rapports de poids soit uni-modale, avec un mode et une moyenne proche de 1. La variation attachée à cette distribution doit de plus être la plus faible possible. Dans de telles conditions, la modification de l'échantillon effectuée lors du calage est répartie de manière uniforme sur la population.

Il est possible d'agir sur la répartition des rapports de poids en modifiant les marges. Ainsi, des secteurs d'activité peuvent être ré-agrégés. Dans plusieurs enquêtes thématiques, le secteur des combustibles et celui de l'eau, du gaz et de l'électricité sont regroupés dans un unique secteur de l'énergie. On peut aussi passer de la NES114 à la NES36. L'enquête n'est alors plus représentative aux mêmes niveaux.

3.3.2 Examens individuels des poids

Il faut s'assurer que les rapports de poids les plus forts ne sont pas attribués à des entreprises atypiques, telles les grandes entreprises.

Lorsque les mesures décrites précédemment ne sont d'aucun effet ou que l'on veut garder le même niveau de précision dans les strates, il est envisageable de caler l'enquête sans ces individus atypiques. Pour plusieurs enquêtes, les marges sont calculées sans les deux grandes firmes de l'industrie automobile –Renault et Peugeot. Le calage se fait alors dans l'industrie automobile sur des entreprises homogènes. Puis Renault et Peugeot sont réintroduits dans l'enquête avec un poids de 1.

3.3.3 La précision des estimations

Les marges de l'échantillon après calage doivent être proches des marges théoriques : l'algorithme de calage de la macro CALMAR donne des résultats calés selon un niveau de précision paramétrable. Le niveau de précision peut être modifié si cela permet d'atteindre de meilleurs rapports de poids. Un calage en milliers d'euros peut ainsi être préféré à celui en euros, si les rapports de poids sont mieux distribués.

³ dans le cas où la non réponse est corrigée préalablement à l'enquête. Dans les autres cas, le coefficient de calage prend en compte la non-réponse et peut être différent de 1.

Après redressement, il est aussi utile de vérifier que d'autres variables intéressantes non utilisées dans le calage ne sont pas trop éloignées de leurs valeurs connues par ailleurs.

3.4 Ne pas forcer les calages

Le calage ne permet pas de résoudre tous les problèmes qui peuvent survenir lors d'une enquête, notamment les erreurs de mesure.

L'enquête sur le financement de l'innovation montre une diminution du taux d'entreprises innovantes par rapport aux enquêtes précédentes.

Plusieurs portes ont été fermées. Cette diminution du taux d'innovantes ne provient pas d'un biais de non-réponse ou d'un défaut d'échantillonnage. Elle correspond à une volonté des entreprises de ne pas se déclarer innovantes, soit en raison de la difficulté du questionnaire, soit d'autres phénomènes difficiles à quantifier.

Face à cette erreur de mesure, aucun calage n'a été mis en œuvre pour le corriger.

Le problème n'a pas été résolu, mais le redressement a incité à entreprendre des analyses complémentaires. Dans l'enquête sur le financement de l'innovation, deux populations d'entreprises sont distinguées : d'une part celles qui ont bien répondu à l'enquête et qui bénéficient de financements publics à l'innovation ; d'autre part les firmes sous-estimées dans l'enquête qui sont marginalement innovantes, innover ponctuellement et ont peu recours aux aides.

4. Illustration par l'exemple du calage de l'enquête EIIG.

1.1 Résultats du traitement de la non-réponse⁴

Compte tenu de la nature de la non-réponse dans l'enquête EIIG, une étape préalable de correction de cette dernière est nécessaire. Elle conduit à un ensemble de pondérations corrigeant la non-réponse qui varient entre 1 et 2,2 (tableau 1).

Tableau 1 : répartition des coefficients correcteurs de la non-réponse.

Quantile	Estimation
100% Max	2.2
99%	1.7
95%	1.7
90%	1.5
75% Q3	1.5
50% Médiane	1.3
25% Q1	1.1
10%	1.1
5%	1.1
1%	1.0
0% Min	1.0

⁴ Voir annexe 2 pour une description plus détaillée du traitement de la non-réponse grâce aux groupements de réponses homogènes.

Les coefficients supérieurs à 1,7 sont rares et correspondent à de petites entreprises du secteur industriel où les groupes de réponse sont caractérisés par une forte non-réponse.

Ces coefficients sont ensuite réintroduits au niveau du calage. A l'issue du calage, les rapports de poids comparent les pondérations des entreprises après calage aux pondérations après traitement de la non-réponse.

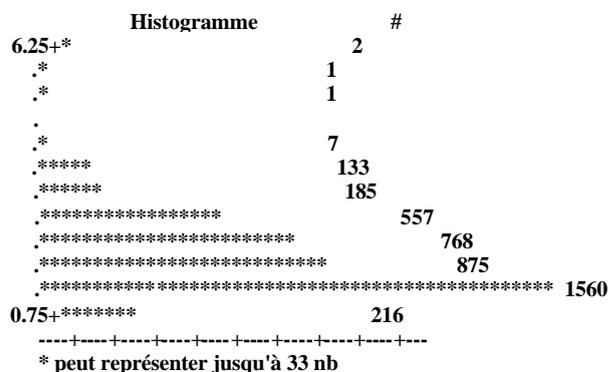
4.2 La mise en œuvre du calage

Les variables de calage retenues sont les montants des flux échangés par les entreprises par secteur (NES36 avec un regroupement des secteur G1 et G2), ainsi que le montant total des échanges de produits pour certaines grandes zones d'échanges (l'Union européenne, l'Amérique du Nord et l'Asie).

La dimension du nombre d'entreprises est abandonnée afin de permettre un meilleur calage. Mais le cadrage en nombre d'entreprises est relativement de bonne qualité. En effet, l'étape préliminaire de traitement de la non-réponse cale chaque groupe de réponses homogènes en nombre d'entreprises.

Les rapports de poids varient entre 0,9 et 2,7. Et les poids finaux varient entre 0,8 et 6 avec des poids proches de 1 pour les entreprises ayant les flux les plus importants et les pondérations les plus fortes pour des entreprises ordinaires de faibles flux, de caractéristiques proches de celles des non-répondantes.

Graphique 1: Répartition des poids de calage.



Les marges de l'échantillon après calage sont satisfaisantes par secteur et par zone d'échange (*tableau 2 et 3*).

Ce calage est de bonne qualité quant aux montants des importations et des exportations de produits (NES36) ainsi que pour décrire les échanges des groupes français et étrangers, bien que ces dimensions n'aient pas servi au redressement.

Tableau 2 : Amélioration des marges après calage en secteur et produits.

Nomenclature NES36	Taux de couverture après calage des flux échangés par secteur en NES36	Taux de couverture après calage des flux échangés par produit, tout secteur confondus, en NES36
B0	100%	99%
C1	95%	111%
C2	100%	90%
C3	100%	97%
C4	100%	99%
D0	98%	97%
E1	97%	99%
E2	99%	101%
E3	100%	97%
F1	100%	99%
F2	96%	92%
F3	100%	104%
F4	99%	100%
F5	99%	101%
F6	100%	98%
G1	100%	100%
G2	100%	100%
J1	94%	99%
J2	100%	80%
Total:	99%	99%

Tableau 3 : Amélioration des marges après calage par zones d'échanges.

Zone d'échanges des biens	Taux de couverture, après calage, par zone d'échange
<i>Union européenne</i>	101%
Europe de l'Est (PECO-CEI)	93%
Autre Europe	93%
Afrique du Nord	90%
Afrique sub-saharienne et australe	81%
Proche et Moyen-Orient	85%
ALENA (USA, Canada, Mexique)	101%
Amérique centrale et du sud (hormis le Mexique)	86%
Japon	111%
Asie hors Japon	96%
Divers pays	76%
Total	99%

5. Conclusion.

Le calage sur marge présente de nombreux avantages pour le redressement d'enquête thématique auprès d'entreprises : meilleure précision, correction possible sous certaines conditions des biais de non réponse et calage sur des grandeurs macroéconomiques.

Le calage permet de gagner en confiance vis à vis des résultats publiés et indique les limites de l'interprétation.

Sa mise en œuvre demande cependant un investissement important. En effet, les étapes préparatoires à l'utilisation de la procédure sont contraignantes (mise en forme des tables en entrée, calcul des marges ...). Mais ensuite, les méthodes et programmes mis en place sont reproductibles d'une enquête à l'autre.

Il est cependant à regretter que les calculs de gains de précision et de diminution de biais obtenus par l'utilisation de Calmar ne soient pas donnés directement parmi les sorties standard de la procédure : cela constituerait un critère supplémentaire d'appréciation de la qualité du calage.

Enfin, le calage ne permet pas de résoudre deux types de problème. D'une part il ne permet pas de corriger les biais trop importants. D'autre part il ne peut pas non plus rectifier les erreurs de mesure. Ainsi, le décalage entre unité enquêtée et celle répondante dépasse le seul cadre du calage sur marges. C'est par exemple le cas de plus en plus fréquent de réponses faites par une entreprise, qui au lieu de répondre pour elle-même le fait pour l'ensemble du groupe.

Bibliographie.

- [1] CARON N., "Les principales techniques de correction de la non-réponse, et les modèles associés", *Documents de Travail 'Méthodologie Statistique n°9604*.
- [2] CARON N., "Application des méthodes de calage à l'enquête EAE-Commerce", *Documents de Travail Direction des Statistiques d'Entreprises n°E2002/05*.
- [3] GUANNEL B., MABILE S., PLATEAU C., "Les échanges internationaux au sein des groupes: une focalisation sur quelques produits et vers l'Union européenne", *4 pages du Sessi n°167*, novembre 2002.
- [4] GUANNEL B., "Traitement de la non-réponse et redressement de l'enquête Echange Internationaux Intra groupe 1999 (EIIG)", *Note de la SDE référence SDE - 16/2002/bg*, septembre 2002.
- [5] LHOMME Y., "Enquête sur les services : Calcul de coefficients de redressement selon la méthode Calmar", *Note de la SDE référence SDE-16bis/2001/YL*, février 2001
- [6] LHOMME Y., "Redressements et analyse de la qualité des réponses de l'enquête thématique FIT", *Note de la SDE référence SDE-17/2001/YL*, mars 2001
- [7] LHOMME Y., "Biais et non-réponse de l'enquête thématique sur le financement de l'innovation", *Note de la SDE référence SDE-42bis/2001/YL*, novembre 2001.
- [8] LHOMME Y., "Première exploitation des résultats de l'enquête CIS3 en vue d'un diagnostic qualité.", *Note de la SDE référence SDE-01/2002/YL*, janvier 2002.
- [9] LHOMME Y., "Qualité de l'enquête communautaire sur l'innovation CIS3", *Note de la SDE référence SDE-22/2002/YL*, mai 2002.
- [10] RANCOURT E., "La non-réponse dans les enquêtes et recensements", *formation continue diplômante de l'attaché*, ENSAI, décembre 2000.
- [11] SAUTORY O., "La Macro CALMAR", *formation continue diplômante de l'attaché*, ENSAI, 2001.
- [12] Sessi, "Le guide des enquêtes thématiques", *Document de travail du Sessi*, août 2002

Annexe 1 : Les caractéristiques des enquêtes Fit, EIIG et CIS3.

L'enquête sur le Financement de l'Innovation Technologique (Fit)

L'enquête Fit porte sur la période 1997-1999. Elle identifie les modes de financement utilisés par les entreprises de l'industrie manufacturière pour financer leurs projets innovants. Elle interroge aussi ces industries sur leurs attentes et souhaits en matière de sources de financement.

Cette enquête porte sur les entreprises industrielles de 20 salariés et plus, à l'exclusion des industries agroalimentaires et du BTP. 5500 entreprises ont été enquêtées. 70% des industries interrogées ont répondu à l'enquête.

Tableau 4 : Répartition du taux de réponse par tranche d'effectif.

Population enquêtée	Taux de réponse dans FIT
Hors tranches	27.3
20 à 49 salariés	77.4
50 à 99 salariés	70.4
100 à 249 salariés	69.7
250 à 499 salariés	65.2
500 salariés et plus	61.1
500 à 999 salariés	60.5
1000 à 1999 salariés	64.0
2000 salariés et plus	58.8
Total	70.1

L'enquête Fit a une répartition non uniforme de la non-réponse (*tableau 4*) : le taux de réponse se dégrade avec la taille de l'entreprise et le minimum de ce taux de réponse est atteint pour la strate des entreprises des 500 salariés et plus (61%).

Il existe un écart entre le taux de réponse en nombre d'entreprises et le taux de couverture des entreprises répondantes en terme de chiffres d'affaire. 70% des entreprises enquêtées sont répondantes, ce qui est moyennement bon, mais elles ne représentent qu'environ 60% du chiffre d'affaire total.

Enfin, le taux d'entreprises innovantes de l'enquête Fit diminue par rapport à celui de l'enquête communautaire sur l'innovation (CIS2,1994-1996).

Cette baisse est particulièrement forte pour les strates des 50 à 99 salariés et celle des 500 à 999 salariés. La baisse du taux d'entreprises innovantes résulte vraisemblablement d'une erreur de mesure.

L'enquête Echanges Internationaux Intra Groupe (EIIG)

L'enquête EIIG a été lancée en 2000 et porte sur la situation en 1999 des groupes internationaux industriels implantés en France.

Elle concerne les entreprises industrielles ou de commerce de gros, implantées en France, dont la somme des importations et des exportations est supérieure à 1 million d'euros et qui appartiennent à un groupe international industriel, c'est à dire à un groupe possédant au moins une filiale industrielle et une filiale à l'étranger.

Son objectif est l'analyse, quantitativement et qualitativement, des flux internationaux de biens manufacturés entre les entreprises situées en France de groupes internationaux industriels, avec les entreprises de ce même groupe implantées à l'étranger. En second lieu, elle apporte un éclairage sur les stratégies des entreprises, voire des groupes, qui s'organisent au niveau mondial.

A l'origine, il s'agissait d'une enquête exhaustive portant sur les 8236 entreprises du champ d'enquête théorique. Mais le champ de l'enquête n'a pu être complètement défini lors du lancement de l'enquête : seulement 5756 entreprises ont été interrogées dont 4305 sont répondantes. Ainsi par rapport au 8236 entreprises devant être théoriquement enquêtées, seules 52% sont répondantes au final (tableau 5). Ces 4305 entreprises répondantes couvrent tout de même 78% des flux d'importations et d'exportations du champ théorique d'enquête. Les entreprises ayant les flux les plus importants ont donc été enquêtées. Celles qui n'ont pas été interrogées ont des flux peu importants.

Tableau 5: Répartition de la réponse suivant le nombre d'entreprises et le montant de leurs échanges douaniers de biens manufacturés, dans l'enquête EIIG.

	Nombre d'entreprises	Montant des Flux en Milliards d'Euros	En pourcentage du nombre théorique d'entreprises	En pourcentage du montant théorique des flux
Champ théorique d'enquête	8236	374	100%	100%
Entreprises enquêtées	5756	352	70%	94%
Entreprises enquêtées répondantes	4305	292	52%	78%

L'enquête EIIG fait face à deux types de « non-réponse » :

- ◆ D'abord, par rapport aux 8236 entreprises du champ théorique d'enquête qui auraient dû être interrogées, 2480 ne l'ont pas été car non repérées comme faisant partie au moment du lancement du champ de l'enquête. Leurs réponses font donc défaut. Le biais engendré par ce premier type de « non-réponse ». doit être corrigé avant redressement.
- ◆ Ensuite, parmi les entreprises enquêtées, certaines sont « non-répondantes » au sens classique du terme : elles n'ont pas pu ou voulu répondre au questionnaire. Ce second type de non-réponse doit aussi être pris en compte dans le traitement du biais de non-réponse.

Les causes de ces deux phénomènes étant très différentes, tout comme les caractéristiques des entreprises concernées, il a fallu réaliser une étape spécifique de traitement de chacun des types de non-réponse. Elle précède le calage sur marge, afin de réaliser ce dernier dans de bonnes conditions. Le traitement particulier de la non-réponse a été effectué en utilisant la méthode des Groupements de Réponses Homogènes (GRH, cf. annexe 2).

Cette enquête fait intervenir deux voire trois dimensions d'interprétation : le niveau entreprise (nombre d'entreprises enquêtées), le niveau flux (montant des flux enquêtés) et éventuellement le niveau groupe. L'enquête EIIG est théoriquement exhaustive sur les entreprises enquêtées (c'est à dire que toutes les entreprises du champ d'enquête sont interrogées) mais pour des raisons de taille des questionnaires d'enquête, une entreprise n'est enquêtée que sur une partie des flux d'importation et d'exportation. Ainsi, en moyenne, une entreprise est interrogée sur 92% du montant total de ses échanges douaniers en 1999⁵.

Puisqu'une entreprise peut être considérée comme un ensemble de flux, la dimension « flux » est par conséquent incluse dans la dimension « entreprise ». Ces deux dimensions restent tout de même

⁵ Les 8 % de flux non enquêtés ont été complétés par imputation partielle.

différentes, dans la mesure où un calage au niveau de l'entreprise n'assure pas automatiquement le calage au niveau des flux.

L'enquête communautaire sur l'innovation (CIS3)

L'enquête CIS3 est une enquête communautaire sur l'innovation, lancée en France en 2001 et portant sur la période 1998-2000.

Il s'agit de la quatrième enquête sur l'innovation réalisée en France. Les entreprises sont interrogées sur la nature et la conduite de leurs innovations, les dépenses de recherche et de développement, le chiffre d'affaire en produits innovants, les sources internes et externes de financement, les coopérations pour innover, les objectifs de l'innovation.

5000 entreprises de 20 salariés et plus et qui appartiennent à l'industrie manufacturière ont été interrogées. Le taux de réponse s'élève à 86% pour le nombre d'entreprises, à 89% pour les effectifs et à 89% en chiffre d'affaire (*tableau 6*).

Le taux de réponse est élevé (86%) et la répartition est relativement homogène. Ces deux critères ont été recherchés au cours de la collecte de l'enquête grâce à un "tableau de bord" hebdomadaire qui a permis une relance ciblée des entreprises non-répondantes. Ainsi, s'est-on placé dans une situation favorable pour le calage.

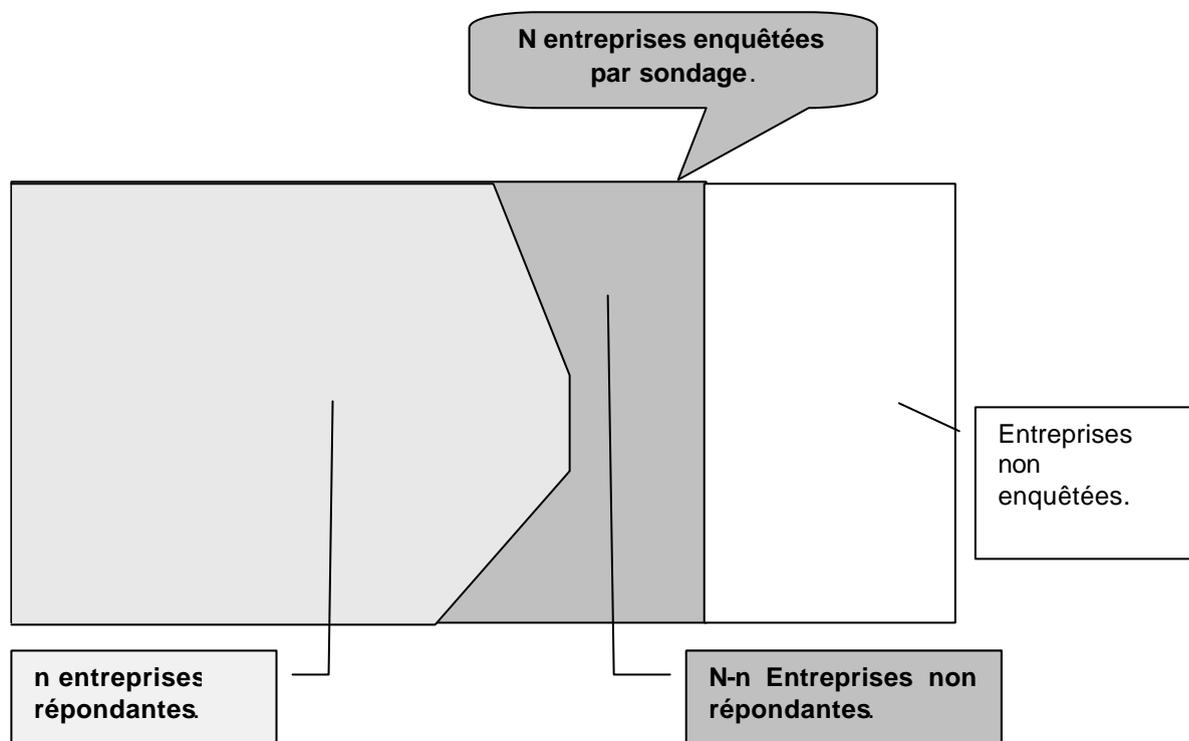
Annexe 2 : Détails du traitement de la non-réponse dans l'enquête EIIG.

Description de la méthode des GRH

Il s'agit d'une méthode de traitement de la non-réponse qui consiste à corriger le biais de non-réponse en déterminant un coefficient de re pondération.

L'idée est de modéliser la non-réponse comme une phase de sondage supplémentaire au sondage initial.

Graphique 2 : les deux types de non-réponse



Ainsi, si on note r le nombre d'entreprises répondantes parmi l'échantillon de départ, un estimateur sans biais du total d'une variable d'intérêt est :

$$\hat{Y} = \sum_{k \in U} \frac{Y_k}{p_k * P(k \in \text{Répondant})}$$

Pour modéliser $P(K \in \text{Répondant})$ on suppose que la non-réponse n'est pas répartie aléatoirement dans le modèle mais qu'il s'agit plutôt d'une répartition en sous-populations supposées homogènes au sens de la non-réponse. Aussi, dans chaque sous-population h on attribue à chaque entreprises la même probabilité de répondre $P(K \in \text{Répondant}) = r_h/n_h$ avec n_h le nombre total d'éléments (répondants ou non) de notre sous-population h et r_h nombre d'éléments répondants de la sous-population h .

Le modèle logistique caractérisant les entreprises du champ théorique non enquêtées.

Pour déterminer les variables expliquant la sélection ou non d'une entreprise dans le champ de d'enquête à son lancement, un modèle logistique est utilisé. Après sélection des variables, les variables explicatives mentionnées dans le *tableau 8* ci-après ont été finalement retenues.

Ce modèle est un modèle logistique dont la variable à expliquer est l'indicatrice de présence d'une entreprise du véritable champ théorique d'enquête, dans les entreprises effectivement enquêtées. Elle vaut 1 si l'entreprise est enquêtée, 0 sinon.

Les variables explicatives testées sont les suivantes :

- Le niveau des flux douaniers de l'entreprise. Cette variable détermine si les flux échangés par une entreprise donnée sont forts, faibles ou moyens, prend 5 modalités :
 - ◆ *Flux les plus forts* : Les entreprises dont le montant des flux imports+exports est parmi les plus forts du champ théorique d'enquête. (supérieurs à 5 milliards de Francs)
 - ◆ *Flux forts* : Les entreprises dont le montant des flux d'imports+exports échangés est élevé sans être parmi les plus forts. (entre 1 et 5 milliards de Francs)
 - ◆ *Flux moyen* : Les entreprises dont le montant des flux échangés est intermédiaire (compris entre 500 millions et 1 milliards de Francs)
 - ◆ *Flux faibles* : Les entreprises dont le montant des flux d'imports+exports est compris entre 100 et 500 millions de Francs.
 - ◆ *Flux les plus faibles* : Les entreprises dont le montant des flux d'imports+exports est parmi les plus faibles (inférieurs à 100 millions de Francs)

Ces groupes ont été construits, en cherchant des regroupements cohérents⁶ d'entreprises, proches par le montant de leurs flux échangés.

On s'attend à ce que plus les flux échangés par une entreprise sont importants, plus cette entreprise est facilement repérable et donc enquêtée. Et cela est d'autant plus vrai que des entreprises qui avaient des montants de flux très importants on été rajoutées au champ de l'enquête même si on n'était pas totalement certain qu'elles appartenaient au champ théorique d'enquête : ces entreprises formaient le *filet de sécurité*.

- Secteur d'activité de l'entreprise. Il s'agit du code d'Activité Principale Exercée (APE) de l'entreprise dans la nomenclature NES16.
- Niveau de visibilité de l'entreprise. Il s'agit d'un indicateur de sollicitation des entreprises : l'idée sous-jacente étant que plus une entreprise est sollicitée lors de différentes enquêtes, plus elle est facile à repérer et donc plus elle a de chance d'avoir été enquêtée. Cette variable est obtenue en comptant le nombre de fois qu'une entreprise apparaît dans les 5 fichiers suivants : fichier groupe du SESSI, fichier LIFI de l'INSEE, les EAE, le fichier des comptes consolidés de la Banque de France et l'enquête des implantations françaises à l'étranger de la France de 1998. Cette variable a donc 5 modalités : par exemple la modalité "2" signifie que l'entreprise se trouve dans 2 des 5 fichiers précédents.
- Nationalité du groupe de l'entreprise. Cette variable est une indicatrice qui vaut 1 si l'entreprise appartient à un groupe de nationalité française (i.e. la tête de groupe de l'entreprise est française) et qui vaut 0 sinon.

⁶ On dit qu'un groupe d'entreprises est cohérent si le *box plot* des flux échangés par ces entreprises n'a pas de valeurs aberrantes ou éloignées.

- Pays d'implantation de l'entreprise. Certaines entreprises ne sont en fait que des établissements basés en France d'entreprises dont le siège social est situé à l'étranger. Ainsi la variable «pays d'implantation» vaut 1 si le pays d'implantation de l'entreprise est la France, 0 sinon. On s'attend à ce que les entreprises situées à l'étranger soient moins enquêtées - car atypiques- que celles implantées en France.

La méthode retenue pour construire le modèle est de commencer par retenir une modalité de référence⁷ pour chacune des variables explicatives choisies. Pour les variables indicatrices, la modalité de référence est 0 par défaut. Pour les autres variables, on choisit la modalité pour laquelle le taux d'entreprise enquêtées est le plus proche du taux d'entreprises enquêtées total (à savoir 70%). De sorte qu'à chaque fois, on raisonne par rapport au taux d'enquête moyen. Ainsi, pour la variable «niveau de flux douaniers» la variable de référence est la modalité *flux moyens*, pour la variable «secteur d'activité» la modalité de référence est le secteur *EF* et pour la variable «niveau de visibilité» la modalité de référence est *un*.

Puis on regroupe entre elles les modalités similaires d'une même variable. L'idée étant que les modalités regroupées ne forment en fait qu'une, car elles contribuent identiquement dans le modèle. On peut aussi enlever certaines modalités du modèle en les regroupant à la modalité de référence.

Tableau 6 : Description du modèle LOGIT, modélisant le fait d'enquêter ou non une entreprise.

Analyse des estimations des maximums de vraisemblance					
Paramètre	DL	Estimation	Erreur standard	Wald Chi-Square	Pr > ChiSq
Constante	1	-1.8330	0.4833	14.3824	0.0001
Niveau des flux douaniers de l'entreprise.					
<i>Flux faibles</i>	1	-2.1208	0.3383	39.2965	<.0001
<i>Flux les plus faibles</i>	1	-3.3514	0.3327	101.4882	<.0001
Secteur d'activité de l'entreprise.					
<i>Secteur des IAA (EB)</i>	1	1.2661	0.1173	116.5300	<.0001
<i>Secteur du commerce de gros (EJ)</i>	1	0.5754	0.0733	61.5638	<.0001
Niveau de visibilité de l'entreprise.					
<i>Deux</i>	1	0.5160	0.1398	13.6278	0.0002
<i>Trois</i>	1	1.6408	0.1386	140.1417	<.0001
<i>Quatre</i>	1	2.8382	0.1399	411.6135	<.0001
<i>Cinq</i>	1	4.3429	0.1714	642.1869	<.0001
Nationalité du groupe de l'entreprise.	1	-0.7956	0.0638	155.5076	<.0001
Pays d'implantation de l'entreprise.	1	3.5670	0.4996	50.9728	<.0001

Les variables utilisées sont bien statistiquement significatives, ainsi que leurs modalités. En définitive, les modalités *moyen*, *fort* et *les plus forts* de la variable «niveau de flux» sont mis en référence et s'opposent au flux faibles avec une gradation dans les flux. De même, pour la variable «secteur d'activité» les secteurs industriels traditionnels s'opposent aux activités moins industrielles comme l'agroalimentaire et le commerce de gros.

⁷ La modalité de référence est la modalité par rapport à laquelle on se base pour savoir si les autres modalités de la variable expliquent mieux ou moins bien le phénomène à expliquer. Elle n'apparaît pas explicitement dans le modèle mais est constamment présente dans l'interprétation d'impact de chacune des modalités sur le modèle.

Ainsi plus une entreprise a des montants de flux d'échanges importants plus elle a de chance d'être enquêtée. Les entreprises des IAA et du commerce de gros ont moins été oubliées que celles de l'industrie (compte tenu de l'important filet de sécurité consenti pour le secteur EJ et du petit nombre d'entreprises du secteur des IAA). Comme on s'y attendait, plus une entreprise est visible plus elle a de chance d'être enquêtée. Les entreprises de groupes étrangers ont plus été enquêtées que les entreprises françaises. Ceci s'explique par l'attention particulière dont elles ont fait l'objet lors de l'enquête. Enfin, les entreprises implantées à l'étranger sont plus difficilement repérables que les entreprises d'implantation française.

Le modèle logistique caractérisant les entreprises enquêtées mais non-répondantes.

Pour corriger les effets de la non-réponse des entreprises enquêtées, la méthode des GRH est utilisée, comme pour le traitement de la non-exhaustivité du champ de la partie précédente. On remplace ici le coefficient n/N par m/M , où M est le nombre d'entreprises enquêtées, et m représente le nombre d'entreprises répondantes parmi les M enquêtées.

Le modèle utilisé, pour expliquer le phénomène de non-réponse, est un modèle logistique : la variable à expliquer est modélisée par une indicatrice valant 1 si l'entreprise a répondu, 0 sinon.

Les variables testées pour expliquer le comportement de non-réponse des entreprises sont les suivantes :

- Le niveau des flux douaniers de l'entreprise. Cette variable est celle utilisée pour le modèle d'explication de la non-exhaustivité de l'enquête, elle est décrite dans la partie précédente de ce document.
On s'attend à ce que les entreprises qui ont les flux douaniers les plus importants soient celles qui répondent le plus au questionnaire, dans la mesure où elles ont fait l'objet de davantage de relances, lors de l'enquête, que les petites.
- Secteur d'activité de l'entreprise. Il s'agit du code d'Activité Principale Exercée (APE) de l'entreprise dans la nomenclature NES16.
- Niveau de visibilité de l'entreprise. Il s'agit de la variable décrite dans la partie précédente. Dans ce modèle, cette variable est utilisée avec l'idée que les entreprises les plus visibles sont celles qui sont le plus sollicitées pour répondre à des enquêtes (une entreprise dont le niveau de visibilité est de 5 est concernée par les enquêtes de 4 institutions administratives). Elles devraient, par conséquent, être plus habituées aux enquêtes que les moins sollicitées et donc répondre plus massivement.
- Nationalité du groupe de l'entreprise. Cette variable est une indicatrice qui vaut 1 si l'entreprise appartient à un groupe de nationalité française (i.e. la tête de groupe de l'entreprise est française) et qui vaut 0 sinon.
- Mode d'enquête. Il s'agit d'une variable indicatrice qui vaut 1 si l'entreprise a été enquêtée par voie postale et 0 si l'enquête a été faite par un enquêteur INSEE directement sur place dans l'entreprise. Les enquêtes faites par enquêteur INSEE étant les plus lourdes (le nombre de questions était plus important que d'habitude) on s'attend à ce que les entreprises soumises à une enquête plus lourde aient moins bien répondu que le reste des entreprises.

Le *modus operandi* est le même que celui décrit dans la partie précédente : on commence par tester la validité du modèle et de chaque variable, puis on effectue au sein de chaque variable le regroupement de modalités ayant les mêmes effets sur le modèle.

Les modalités de référence des variables « niveau des flux douaniers » de l'entreprise et « niveau de visibilité » restent respectivement "moyen" et "un". Pour la variable « secteur d'activité » la modalité de référence est le secteur ED. La modalité de référence pour les indicatrices est encore 0.

Les modalités *forts* et *faibles* de la variable « Niveau des flux » sont regroupées à la modalité de référence *moyen*. La modalité *deux* de la variable « Niveau de visibilité » est regroupée avec la modalité de référence *un*. Au final on obtient le modèle ci-dessous :

Tableau 7 : Description des coefficients sélectionnés grâce au modèle LOGIT.

Analyse des estimations des maximums de vraisemblance					
Paramètres	DL	Estimations	Erreur standard	Wald Chi-Square	Pr > ChiSq
Constante	1	0.0344	0.2210	0.0242	0.8764
Niveau des flux douaniers de l'entreprise.					
<i>Flux très forts</i>	1	0.7549	0.1774	18.1082	<.0001
<i>Flux parmi les plus faibles</i>	1	-0.4001	0.0699	32.8080	<.0001
Secteur d'activité de l'entreprise.					
<i>Secteur des IAA (EB)</i>	1	1.3874	0.1431	93.9616	<.0001
<i>Secteur des biens de consommation (EC)</i>	1	-0.2511	0.0895	7.8713	0.0050
<i>Secteur de l'énergie (EG)</i>	1	-0.8816	0.3735	5.5720	0.0182
<i>Secteur du commerce de gros (EJ)</i>	1	1.1240	0.1048	115.1169	<.0001
Niveau de visibilité de l'entreprise.					
<i>Trois</i>	1	-0.2890	0.1490	3.7597	0.0525
<i>Quatre</i>	1	0.3923	0.1487	6.9607	0.0083
<i>Cinq</i>	1	1.0241	0.1705	36.0686	<.0001
Nationalité du groupe de l'entreprise.	1	-0.5688	0.0718	62.7875	<.0001
Mode d'enquête	1	0.8712	0.1559	31.2165	<.0001

Ainsi on peut dresser la typologie suivante des entreprises répondantes : toutes choses égales par ailleurs, les entreprises qui répondent le mieux sont celles qui sont les plus visibles (celles dont le niveau de visibilité est supérieur à 3) et qui ont des flux douaniers importants. Les entreprises interrogées par courrier sont aussi plus répondantes que celles enquêtées par un enquêteur : ceci provient du fait que les entreprises non enquêtées par voie postale étaient les plus problématiques (enquêtes les plus longues, enquêtes menées dans une nomenclature très agrégée compte tenu de la multitude des produits échangés par l'entreprise). Les entreprises étrangères répondent mieux que les entreprises françaises dans la mesure où elles ont fait l'objet d'une attention particulière (pour permettre l'exploitation de l'enquête sous le prisme de la nationalité du groupe). Enfin, comme pour le modèle précédent d'explication de la non-exhaustivité de l'enquête, le secteur industriel a moins bien répondu que les IAA et le commerce de gros, et la non-réponse est plus prononcée pour les secteurs EC et EG.