

## Méthodologie de l'enquête sur les chaînes d'activité mondiales (CAM)

L'enquête « chaînes d'activité mondiales » cherche à mettre en évidence le phénomène de fragmentation des processus de production mis en œuvre par les sociétés et qui les conduit à externaliser en France ou à délocaliser des tâches ou travaux réalisés jusque là en interne.

### Présentation et historique de l'enquête CAM

L'enquête « Chaînes d'activité mondiales », réalisée en 2012, s'inscrit dans le cadre d'un projet Européen, coordonné par Eurostat. Quatorze autres pays, plutôt de petite taille, ont également effectué une enquête comparable en 2012. L'appellation européenne officielle de l'enquête est « IS-GVC Survey » (International Sourcing and Global Value Chains Survey).

L'enquête CAM cherche, d'une part, à effectuer une « photographie » fin 2011 des chaînes d'activité mondiales des sociétés implantées en France, et, d'autre part, à capter les événements ayant affecté les chaînes d'activités (externalisations en France, délocalisations) de ces dernières entre début 2009 et fin 2011.

Les principaux thèmes abordés portent sur :

- le **type des activités externalisées en France ou délocalisées entre 2009 et 2011** : cœur de métier d'une part, fonctions support découpées en six modalités d'autre part ;
- la **forme des délocalisations entre 2009 et 2011** : au sein du groupe, en dehors de groupe ;
- la **destination des délocalisations entre 2009 et 2011** : découpage de l'étranger en 11 zones géographiques ;
- l'impact des délocalisations en termes de **suppressions de postes en France entre 2009 et 2011** ;
- les **motivations** et les **obstacles** rencontrés lorsqu'une délocalisation a été envisagée ou effectuée **entre 2009 et 2011** ;
- le **recours des sociétés françaises à leurs propres filiales** en France et à l'étranger, détaillé **par activité et zone géographique, fin 2011** ; l'évolution de ce recours à l'étranger **entre 2009 et 2011** ;
- le **recours des sociétés françaises à des sociétés du groupe autres que leurs propres filiales**, en France et à l'étranger, détaillé **par activité et zone géographique** ; l'évolution de ce recours à l'étranger **entre 2009 et 2011** ;

- le **recours des sociétés françaises à la sous traitance en dehors du groupe**, en France et à l'étranger, détaillé **par activité et zone géographique** ; l'évolution de ce recours à l'étranger entre 2009 et 2011 ;
- les **effectifs employés au sein de la société fin 2011 par activité**, et l'évolution de ces effectifs entre 2009 et 2011.

---

## Champ de l'enquête

---

### Période de référence

L'enquête CAM, dont la collecte a eu lieu en 2012, porte :

- sur la période 2009-2011 pour les questions relatives aux externalisations en France, aux délocalisations et aux questions relatives à des évolutions ;
- sur l'année 2011 ou la fin de l'année 2011 pour les autres questions.

### Champ

Les unités interrogées sont les sociétés (unités légales identifiées par leur numéro Siren) :

- actives, marchandes et exploitantes au 31/03/2012,
- employant 50 salariés ou plus fin 2008,
- implantées en France (métropole et DOM),
- dont l'activité principale est classée dans les secteurs correspondant aux sections B à N (hormis section K correspondant aux activités financières et d'assurance) en nomenclature Nace Rév. 2.

Les secteurs en Naf Rév. 2 faisant partie du champ de l'enquête sont explicitement :

- l'industrie extractive (section B) ;
- l'industrie manufacturière (section C) ;
- la production et distribution d'électricité, de gaz, de vapeur et d'air conditionné (section D) ;
- la production et distribution d'eau, assainissement, gestion des déchets et la dépollution (section E) ;
- la construction (section F) ;
- le commerce (section G) ;
- les activités de transport et d'entreposage (section H) ;
- l'hébergement et restaurant (section I) ;
- les activités liées à l'information et la communication (section J) ;
- les activités immobilières (section L) ;
- les activités de services spécialisés, scientifiques et techniques (section M) ;
- les activités de services administratifs et de soutien (section N).

Le champ imposé par Eurostat se limite aux sociétés employant 100 salariés ou plus fin 2008. L'extension du champ aux sociétés employant entre 50 et 99 salariés fin 2008 est un choix français. Afin d'assurer la comparabilité avec les autres pays participant à l'enquête dans les publications d'Eurostat, les résultats que l'Insee a fourni à Eurostat se limitent aux sociétés du champ de 100 salariés ou plus fin 2008.

---

## La réalisation de l'enquête

---

### Période de collecte

la collecte a eu lieu du 15 juin au 19 octobre 2012.

### Mode de collecte

L'enquête a été effectuée par internet, avec possibilité pour les sociétés de répondre par voie postale sur un questionnaire papier sur simple demande. Les sociétés ont reçu initialement un courrier avec la lettre de présentation de l'enquête contenant le lien du site internet de collecte ainsi que l'identifiant et le mot de passe permettant de répondre en ligne.

---

## Plan de sondage

---

### Unités interrogées

L'unité statistique interrogée est la société au sens d'unité légale (identifiée par son numéro Siren)

### Base de sondage

La base de sondage est issue du répertoire SIRENE et contient 28 612 unités appartenant au champ de l'enquête.

Cette base de sondage a ensuite été appariée à la base de données du dispositif LIFI 2009 (Liaisons financières, année 2009) de l'Insee, afin de récupérer notamment des informations sur le type d'unités légales (indépendante, filiale, tête de groupe) et sur la possession de filiales françaises ou étrangères.

### Taille de l'échantillon

L'échantillon comporte 8 093 unités.

### Type de sondage

Sondage stratifié avec tirage aléatoire « systématique » dans chaque strate.

La base de sondage est stratifiée par le croisement de deux critères :

- selon le secteur d'activité de l'unité légale (71 modalités). Les modalités ont été établies selon trois critères :
  - o respect des exigences européennes en terme de résultats attendus par secteur ;
  - o réflexion sur des découpages sectoriels plus fins pour l'exploitation française des résultats ;
  - o attention particulière à ce que le nombre d'unités dans les différents secteurs soit *a priori* de taille « raisonnable », ce qui conduit à effectuer des éclatements ou des regroupements de sous-secteurs en cohérence avec les objectifs de l'enquête. Ce critère s'inscrit aussi dans un souci de précision quant aux résultats de l'enquête.
- selon son effectif salarié fin 2008 : 3 tranches (B : entre 50 à 99 salariés, C : 100 à 249 salariés, D : 250 salariés et plus).

Le croisement de ces deux critères conduit à un total de 213 strates.

### Exhaustivité

Les 4 964 unités légales de 250 salariés et plus du champ sont interrogées exhaustivement (71 strates concernées). Ces unités représentent 61,3 % des unités de l'échantillon, et 17,3 % des unités de la population du champ.

### Tirage dans les strates non exhaustives

Dans les strates ne relevant pas de la partie exhaustivement enquêtée (142 strates concernées), une allocation proportionnelle aux effectifs salariés a été calculée, puis modifiée à la marge de façon à améliorer (ou rééquilibrer selon les cas) les précisions sur les domaines de diffusion les plus attendus :

- réduction des taux de sondage dans le secteur du commerce ;
- tirage de 6 unités au minimum dans les strates où le nombre total d'unités est d'au moins 6 ;
- tirage de toutes les unités présentes dans les strates où le nombre d'unités est inférieur à 6.

Pour chaque strate de la partie non exhaustive, le tirage de l'échantillon a été réalisé selon l'algorithme de tirage systématique sur la base de sondage triée au sein de chaque strate par « type fin » d'unité, afin d'assurer une certaine représentativité des différents types d'unités dans chaque strate.

Les 8 modalités qualifiant le « type fin » d'une unité sont les suivantes :

- indépendante ;
- filiale d'un groupe installé uniquement en France ;
- tête d'un groupe installé uniquement en France ;
- tête d'un groupe multinational ;
- filiale d'un groupe multinational, possédant des filiales à l'étranger et avec une tête de groupe française ;

- filiale d'un groupe multinational avec tête de groupe à l'étranger et possédant des filiales à l'étranger ;
- filiale d'un groupe multinational avec tête de groupe à l'étranger et ne possédant pas de filiale à l'étranger ;
- filiale d'un groupe multinational, ne possédant pas de filiale à l'étranger et avec une tête de groupe française.

---

## Redressement de la non-réponse et calage

---

Le taux de réponse global à l'enquête est de 80 %, ce qui implique qu'une partie de l'échantillon (20 %) est en situation de non-réponse totale à l'enquête. Par ailleurs, certaines unités ayant répondu à l'enquête sont en situation de non-réponse partielle. D'autres encore ont pu renseigner certaines réponses incohérentes ou invraisemblables entre elles ou vis-à-vis d'autres sources statistiques. Afin d'obtenir des résultats de meilleure qualité possible, il est donc nécessaire d'effectuer une validation des données, un redressement de la non-réponse totale et partielle et un calage des pondérations.

Ainsi, pendant la collecte, l'équipe de gestion procède à des rappels d'unités enquêtées afin d'éclaircir avec les répondants, par téléphone ou par courriel, une partie des données jugées incohérentes ou invraisemblables. Cette phase est appelée « phase d'expertise des données par l'équipe de gestion ». Cette phase a eu lieu entre juillet et octobre 2012 pour l'enquête CAM.

Après la collecte, des contrôles systématiques et plus complexes sont appliqués aux données pour en éliminer toute incohérence ou invraisemblance. Cette phase est appelée phase d'« apurement des données ». Cette phase a eu lieu entre mi-novembre et fin décembre 2012.

Enfin, la phase d'apurement est suivie par la phase de « redressement de la non-réponse et de calage ». Cette phase englobe la correction de la non-réponse totale (modification des poids des unités afin de garantir que l'exploitation des réponses sera exactement représentative du champ de l'enquête) et partielle. Cette phase comprend également un calage des poids des unités sur des variables de cadrage afin d'améliorer la précision des estimations d'agrégats. Cette phase, qui a eu lieu entre janvier et mi-mars 2013, est détaillée dans la suite.

### Classement des unités en « substituable » ou « non-substituable »

Lors du traitement de la non-réponse, on distingue les unités « substituables » des unités « non-substituables ». Les unités dites « non-substituables » sont, dans l'enquête CAM, celles ayant un effectif salarié fin 2008 au moins égal à 8 000. Ces unités ont un poids économique trop important pour qu'elles soient soumises à une correction de la non-réponse totale par repondération. Par contre, les unités non substituables en situation de non-réponse partielle seront soumises aux mêmes méthodes que les unités substituables dans la même situation. Les unités ayant un effectif salarié fin 2008 inférieur à 8 000 sont dites « substituables ».

### Répartition « répondant / non-répondant / hors-champ »

Afin de distinguer clairement les entreprises hors champ (qui participeront au calage mais pas à la correction de la non-réponse) et les entreprises « non-répondantes totales » (dont les poids sont redistribués sur les répondantes et qui ne participent pas au calage), la première démarche est de coder, pour chaque unité de l'échantillon, une variable « **ETAT** » selon 4 modalités :

- modalité « **R** », unité **répondante du champ** : l'unité a retourné un questionnaire valide et son appartenance au champ est avérée ;
- modalité « **N** », unité **non-répondante du champ** : l'unité n'a pas retourné son questionnaire (ou celui-ci est trop imparfaitement rempli pour être considéré comme valide ;
- modalité « **H** », unité **hors champ** : les informations recueillies permettent de conclure qu'elle n'appartient finalement pas au champ de l'enquête ;
- modalité « **C** », unité **cessée avant le 01/01/2010**.

## Phase de détermination du statut d'une variable ou d'un bloc de variables

Les variables dites de **statut** sont déterminées uniquement pour les unités répondantes (« ETAT = R ») et pour les « non-substituables » non-répondantes.

Ces variables indiquent pour chaque variable ou chaque bloc de variables :

- si l'on dispose d'une réponse valide (valeur « B »),
- s'il n'a pas de réponse alors que l'on en attendait une (valeur « M »),
- s'il n'y a pas de réponse parce que l'unité n'est pas concernée par la question (valeur « N »).

Ces variables permettent :

- d'imputer une réponse uniquement aux variables non renseignées alors que l'entreprise est concernée par la question ;
- de distinguer dans la table finale les réponses fournies directement par l'unité de celles issues de l'apurement ou celles imputées lors du redressement.

## Redressement de la non-réponse partielle

### Recherche de liens entre différentes variables

Dans un premier temps, on recherche des liens pouvant exister entre la variable à redresser et d'autres variables dites auxiliaires. Ces variables peuvent être soit des réponses à d'autres questions de l'enquête, soit des données issues de sources externes à l'enquête tel que le secteur ou l'effectif salarié. La recherche de liens s'effectue grâce aux tests du Khi-Deux ou à une méthode de segmentation par arbre de segmentation reposant sur un algorithme de type CHAID (qui fait appel également à des tests du Khi-deux).

### Imputation des variables qualitatives

La non-réponse à certaines variables qualitatives a fait l'objet de corrections manuelles, après un traitement au cas par cas, notamment en utilisant diverses sources externes. Par exemple, pour les questions de l'enquête portant sur le type de l'unité (appartenance à un groupe, contrôle de filiales, ...), la base de données LIFI 2010 (liaisons financières) a été utilisée.

Les autres variables ont été redressées par une procédure de « hot-deck », c'est à dire par imputation de la réponse à la même question d'une unité valide (c'est à dire avec « statut = B ») et ayant des caractéristiques proches de l'unité à redresser. Ces caractéristiques ont été recherchées parmi des variables présentant un lien avec la variable à redresser. Les imputations effectuées sur les unités de grande taille ont été systématiquement examinées a posteriori.

### Imputation des variables quantitatives (2 questions)

Les imputations des réponses à la question 2.7 sur le nombre de postes supprimés suite aux délocalisations effectuées entre 2009 et 2011 ont été exclusivement manuelles et fondées sur le recoupement de diverses sources externes.

Plusieurs techniques ont été combinées selon les besoins pour imputer les réponses à la question 6.1 sur la répartition de l'effectif salarié par activité fin 2011 (cœur de métier et six fonctions supports) : imputation manuelle, imputation par hot-deck et imputation par la valeur moyenne de groupes d'unités adéquatement formés.

## Redressement de la non-réponse totale

Les unités « non substituables » ne peuvent représenter qu'elles-mêmes, leur poids est donc toujours égal à 1. Elles sont de ce fait exclues du procédé de correction par repondération lors du redressement de la non-réponse totale.

*Remarque : les unités de 250 salariés ou plus interrogées exhaustivement (soit un peu plus de 60 % de l'échantillon) ont toutes un poids de lancement valant 1, mais le poids des unités légales de 250 salariés ou plus « substituables » peut être modifié lors de la correction de la non-réponse totale.*

Pour redresser la non-réponse totale, il a été appliqué la méthode dite de « repondération », à l'aide de groupes de réponses homogènes (GRH). La repondération vise à augmenter le poids des unités « répondantes du champ » pour compenser celui des unités « non-répondantes du champ ».

La méthode consiste à bâtir un modèle explicatif du comportement de réponse à partir de variables auxiliaires et à retenir les variables qui paraissent significatives. Les variables auxiliaires doivent être connues sur toutes les unités du champ, répondantes ou non, ce qui exclut le recours aux questions de l'enquête. Elles doivent donc, soit être présentes dans le fichier de lancement, comme le secteur d'activité, la tranche de taille ou encore la localisation (Paris/Province...), soit être mobilisables à partir de sources externes comme le taux d'exportation ou d'endettement.

Pour constituer les GRH, les variables auxiliaires suivantes ont été utilisées :

- l'état économique de l'unité dans Sirène, en 2 modalités :
  - o active ;
  - o inactive.
- la zone d'implantation de la société, selon 3 regroupements :
  - o Bassin Parisien + Est ;
  - o Nord + Ouest + Sud-ouest + Centre-est ;
  - o DOM + Région Parisienne.
- la tranche d'effectif salarié, en 2 modalités très agrégées :
  - o 50 à 99 salariés
  - o 100 à 249 salariés ou 250 salariés ou plus
- l'activité, en 2 modalités très agrégées :
  - o industries (B : industries extractives ; C : industrie manufacturière) ;
  - o autres activités.
- le taux d'endettement de l'unité (source externe) en 2 modalités :
  - o Moins de 100 % ;
  - o 100 % ou plus.
- le type d'unité en terme organisationnel (source LIFI), en 2 modalités :
  - o n'est pas indépendante et appartient à un groupe franco-français ;
  - o est indépendante ou n'appartient pas à un groupe franco-français.

Les variables significatives ont été déterminées à l'aide d'une régression logistique. Tous les croisements possibles entre ces variables n'ont pas été retenus pour constituer les GRH : des groupes « adjacents » ont été réunis lorsque leur taux de réponse n'était pas significativement différent.

Une fois les GRH constitués, on a procédé à une repondération pour corriger la non-réponse totale. La probabilité de réponse à l'intérieur de chaque GRH a été estimée comme le rapport entre la somme des poids des unités répondantes du GRH et la somme des poids de l'ensemble des unités du GRH. Le poids d'une unité répondante après correction de la non-réponse est obtenu en divisant son poids au lancement par la probabilité de réponse estimée pour le GHR auquel elle appartient. Quant au poids des unités « non-répondantes totales » du champ, il est ramené à zéro (car transféré aux autres unités).

*Remarque : Le poids des unités hors champ n'est pas modifié lors de cette phase.*

## **Calage**

Les méthodes de calage d'un échantillon consistent à changer à la marge les poids obtenus après le traitement de la non-réponse totale, pour que les estimations de totaux de variables quantitatives de cadrage effectuées à partir des pondérations de l'enquête (ces totaux correspondant la plupart du temps à des « nombres » de sociétés selon diverses caractéristiques) soient égales aux totaux connus par ailleurs sur la population.

Les unités participant au calage sont les unités répondantes du champ et les unités hors-champ présentes dans la base de sondage. Les unités de l'échantillon classées en hors-champ sont censées représenter toutes les unités de la base de sondage qui auraient été classées hors-champ si elles avaient été échantillonnées.

L'outil utilisé afin de mettre en œuvre le calage est la macro « CALMAR » codée en langage SAS, (macro disponible sur le site internet de l'INSEE). Cette macro a été utilisée d'une manière très classique, en prenant la méthode dite du « raking ratio » puis celle dite du « logit tronqué » avec des limites inférieure et supérieure pour les rapports des poids choisies relativement proches de 1. Quelques petites classes de calage causant une trop grande distorsion des poids ont été regroupées. Enfin, quelques unités avaient un poids inférieur à 1 à l'issue du calage, leurs poids ont été augmentés et en contrepartie, les poids d'unités dont le poids était le plus grand ont été diminués d'autant.

Les variables retenues pour le calage ont été :

- l'effectif salarié au lancement ;
- le nombre de sociétés par tranche d'effectif (de 50 à 99 salariés, de 100 à 249 salariés, 250 salariés et plus) ;
- le nombre de sociétés par secteur d'activité tel qu'il a été utilisé dans le plan de sondage (le plus souvent au niveau division), avec quelques regroupements de secteurs lorsque cela a été nécessaire ;
- le nombre de sociétés qui contrôlent des filiales à l'étranger ;
- le nombre de sociétés qui contrôlent des filiales en France ;
- le nombre de sociétés selon leur taux d'exportation en pourcentage de leur chiffre d'affaires, en trois classes : moins de 1%, de 1 à 50 % et plus de 50 % ;
- le nombre de filiales contrôlées à l'étranger ;
- le nombre de filiales contrôlées en France ;
- le nombre de sociétés par type d'unité (indépendante, membre d'un groupe français ou international, tête de groupe).