

DIFFUSER DES TAUX D'ÉVOLUTION ENTRE L'ANNÉE N ET N-1, C'EST BIEN, DÉCOMPOSER CETTE ÉVOLUTION EN DIFFÉRENTS FACTEURS EXPLICATIFS, C'EST ENCORE MIEUX !

Heidi KOUMARIANOS

(*) Insee, Département Répertoires, Infrastructure et statistiques structurelles

Introduction

En statistique d'entreprises¹, la majorité des résultats est diffusée par secteur. Le secteur est déterminé selon l'activité principale de l'entreprise (APE), l'information relative à celle-ci étant déclarative ou observée dans une enquête par le biais des différentes activités élémentaires de l'entreprise dites branches d'activité².

Ce classement représente une vision partielle de l'activité de l'entreprise et peut évoluer au cours du temps. Ainsi, les résultats agrégés par secteur sont dépendants des classements individuels et l'évolution d'un agrégat peut recouvrir plusieurs phénomènes.

La diffusion de données agrégées peut se faire principalement sous deux formes :

- des agrégats en niveau : ceux-ci permettent de connaître la valeur des caractéristiques (le nombre d'entreprises, le chiffre d'affaires etc...) pour une année donnée et pour les différents secteurs d'activité ;
- des agrégats en évolution : ils permettent de connaître l'évolution économique d'un secteur pour deux années successives.

Cependant l'explication d'une évolution n'est pas chose aisée. Elle peut masquer des réalités économiques diverses. Ainsi, d'après les résultats Esane, entre 2008 et 2009, le chiffre d'affaires du secteur « Stockage et Entreposage » (521) a baissé de 0,7 %. Dans le même temps, le chiffre d'affaires des entreprises pérennes (présentes dans le secteur en 2008 et 2009) a progressé de 2,8 %. Les entreprises pérennes ne représentent toutefois que les trois quarts du chiffre d'affaires du secteur, et génèrent donc 2,1 points dans l'évolution sectorielle. A l'inverse, les mouvements de champ pèsent pour - 3,5 points (- 8 points pour les entreprises qui ont changé de secteur, et 4,5 points pour celles qui sont entrées dans le secteur). Ainsi l'évolution observée indique plus une modification de classement sectoriel qu'une baisse de l'activité des entreprises qui composent ce secteur. A l'inverse, dans le secteur « Construction de bâtiments résidentiels et non résidentiels » (412) qui a connu une baisse de 18 %, ce sont bien les entreprises pérennes du secteur qui ont vu leur activité diminuer (-18,4 %).

Le classement sectoriel des entreprises étant connu par le biais d'enquête, l'analyse des résultats en évolution s'appuiera nécessairement sur les résultats observés sur l'échantillon de l'enquête structurelle. C'est sur ce point que résident les principales difficultés méthodologiques et de mise en œuvre du calcul des résultats en évolution que nous présenterons par la suite.

L'objectif de ce papier est de présenter les éléments qui permettront d'appréhender au mieux l'évolution d'un secteur entre deux années successives. Plus qu'une mesure classique de l'évolution d'un agrégat entre deux années, il s'agit d'estimer les facteurs expliquant cette dernière en identifiant différentes populations. Après avoir rappelé brièvement le contexte dans lequel se place cette étude, à savoir l'examen des résultats du processus d'Elaboration des statistiques annuelles d'entreprises (ESANE), cet article s'attachera à décrire dans un premier temps les différentes catégories qu'il convient de différencier afin de mettre en évidence les différents facteurs explicatifs de l'évolution. Dans un deuxième temps, les choix méthodologiques qu'il a été nécessaire de prendre, afin de réussir

¹ Par abus de langage, on appellera entreprise dans la suite du document une unité légale.

² Ainsi la société Peugeot par exemple réalise du chiffre d'affaires sur les branches d'activités de « construction automobile », « de commerce de véhicule », « de recherche » et fait partie du secteur « construction automobile » qui reste sa principale activité.

à en estimer le poids économique de chacune de ces sous-populations seront présentés. Enfin nous présenterons quelques résultats, en s'interrogeant sur la fiabilité qu'on peut leur accorder.

1. Quelques rappels et définitions :

1.1. Les sources de données ESANE [1] et [2]

Depuis 2008, la source de données ESANE permet de publier des résultats agrégés de statistique structurelle d'entreprises en utilisant conjointement plusieurs sources de données : des sources administratives « exhaustives » (caractéristiques fiscales et d'emploi) et deux enquêtes statistiques.

La source fiscale donne des informations sur la comptabilité des entreprises sur plus de 3 millions d'unités légales. Environ 600 000 unités considérées comme actives ne figurent pas dans cette source pour différentes raisons, et font l'objet d'imputation individuelle. Deux enquêtes ont remplacé les Enquêtes Annuelles Entreprises (EAE) : l'Enquête Annuelle de Production (EAP) qui concerne l'industrie manufacturière hors IAA et l'industrie énergétique et l'Enquête Sectorielle Annuelle (ESA) pour les autres secteurs. Ces deux enquêtes sont réalisées auprès d'un échantillon de 160 000 entreprises. Leur principale utilité est de collecter la répartition du chiffre d'affaires en branches d'activité élémentaires et d'en déduire l'APE des entreprises. Ces enquêtes sont indispensables pour obtenir des classements sectoriels et donc des agrégats sectoriels fiables.

1.2. Les échantillons des enquêtes

Deux enquêtes couvrent le champ statistique qui nous intéresse. Elles font l'objet de stratégies d'interrogation différentes.

L'EAP couvre le champ des entreprises de l'industrie manufacturière (162 000 entreprises). La composition de son échantillon est la suivante :

- 32 000 entreprises sont interrogées exhaustivement (les seuils d'exhaustivité sont de 20 salariés ou de 5 millions d'euros de chiffre d'affaires).
- 50 000 entreprises sont interrogées par sondage, par rotation d'échantillon, au rythme d'1/6.
- Les 80 000 entreprises les plus petites (effectif nul ou un seul produit pour la NAF concernée) ne sont pas interrogées et sont exclues de la base de sondage. Elles ne sont donc pas représentées par l'échantillon.

L'ESA couvre un champ plus large, qui comprend les secteurs des IAA, de la construction et promotion immobilière, du commerce, des transports et des services. Ce champ comprend plus de 2,5 millions d'unités.

- 56 000 entreprises sont enquêtées exhaustivement.
- 66 000 sont interrogées par sondage. L'échantillon est chaque année renouvelé par moitié (chaque entreprise dispose d'un numéro de tirage et sera conservée si la parité de ce numéro est la même que la parité de l'année).

Les entreprises appartenant à la partie conservée de l'échantillon représenteront les mêmes entreprises (on conserve la strate de tirage et le poids au lancement pour ces unités).

Schéma 1 : Tirage des échantillons de l'ESA et de l'EAP

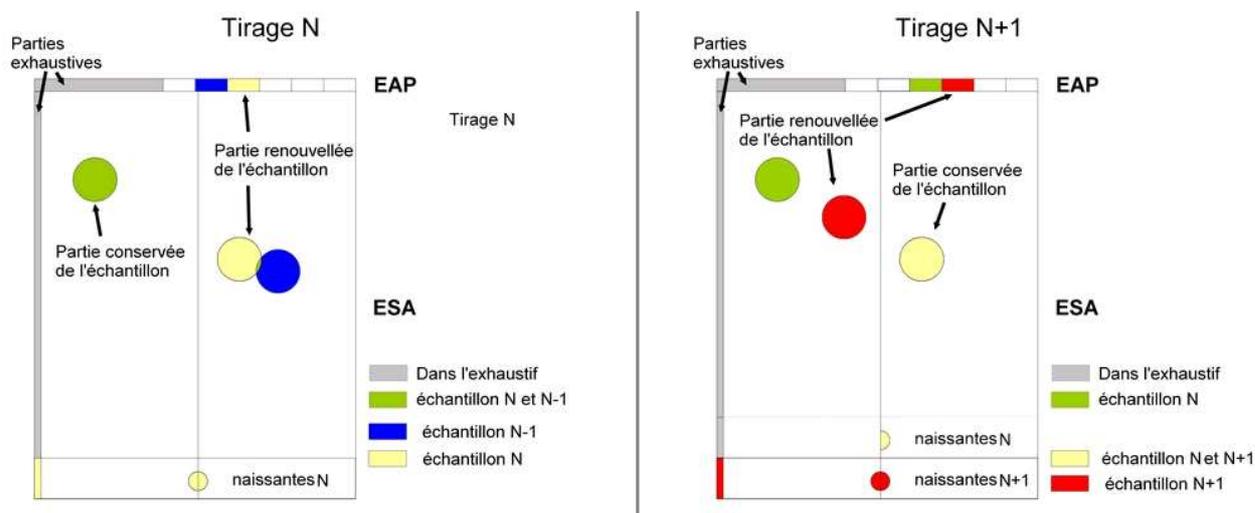


Tableau 1 : Répartition des unités interrogées selon les différentes parties de la base de sondage

Partie de la base de sondage	nombre d'unités de l'échantillon ESA 2010
conservée et échantillonnée	29 582
naissantes et échantillonnée ³	5 760
renouvelée et échantillonnée	30 751
exhaustive	56 285 Dont 734 naissantes
Total	122 378

Source : Insee - tirage de l'échantillon de l'ESA

Pour connaître les changements sectoriels réellement imputables à une année donnée, il est nécessaire de connaître les vrais classements en N-1 et en N, et pour cela, d'interroger certaines entreprises deux années consécutives. Le plan de sondage a été construit à cet effet.

Ce point sera très important par la suite, puisque la connaissance des classements en N et en N-1 sera déterminante pour l'analyse. Ainsi seules les parties exhaustives et conservées de l'échantillon seront retenues pour les différents calculs des résultats en évolution.

Par définition, les entreprises naissantes et interrogées en N représentent l'ensemble des entreprises naissantes en N. Il n'y a pas lieu de poser la question d'une observation en N-1 pour des entreprises naissantes en N. En revanche, en N+1, les entreprises naissantes en N seront conservées ou renouvelées de la même façon que le reste de l'échantillon.

³ les entreprises « naissantes » ne sont pas uniquement des entreprises récemment créées. Ce sont aussi des entreprises réactivées, ou encore des entrées de champ (modification d'une caractéristique de l'entreprise la faisant entrer dans le champ). En 2010, l'échantillon des « naissantes » comprenait 3 800 créations, 450 réactivations et 1 650 entrées de champ.

L'échantillon N est constitué au début du quatrième trimestre N-1, on a donc considéré comme créées les entreprises ayant une date de création en N et en N-1.

1.3. Le traitement des non répondantes dans les enquêtes [3] et [4]

Dans l'EAP, les non répondantes sont extrapolées individuellement.

Dans l'ESA, les non répondantes appartenant à la strate exhaustive sont extrapolées individuellement, soit à partir des données historiques, soit par moyenne de strate en se basant si possible sur le chiffre d'affaires connu dans les liasses fiscales pour l'unité. La non-réponse pour la partie échantillonnée de l'enquête est effectuée par repondération.

Il y a ensuite une étape de winsorisation (troncature de poids élevés), puis un calage sur les marges de chiffre d'affaires et de nombre d'entreprises.

Suite à ces différentes étapes, le rapport des poids initiaux et des poids finaux varie autour de 1,6.

Tableau 2 : Distribution des rapports de poids avant et après calage des entreprises répondantes

	Minimum	1er quartile	Médiane	Moyenne	3ème quartile	Maximum
2009	0,009	1,328	1,539	1,652	1,886	13,895
2010	0,058	1,311	1,492	1,624	1,776	173,230

Ainsi, les poids après calage peuvent différer d'une année sur l'autre, et ce même pour les entreprises appartenant à la partie conservée de l'échantillon.

Le poids dans l'évolution d'une entreprise pourra donc être lié non seulement à l'évolution réelle de ses caractéristiques, mais également à la variation de son poids de sondage d'une année sur l'autre.

Ce point sera discuté dans la partie 3.

1.4. Les restructurations d'entreprises [5]

Une restructuration économique se caractérise par un ensemble d'opérations qui a un impact sur l'activité productive courante (opérations de production et de formation brute de capital fixe) des entreprises par le biais de transferts d'activités.

Les restructurations les plus connues sont les fusions, les absorptions et les scissions partielles ou totales. Elles ont un impact non négligeable sur l'évolution entre deux années consécutives des agrégats sectoriels.

Prenons l'exemple d'un secteur dans lequel toutes les unités voient leur chiffre d'affaires croître de 10% entre deux années consécutives et qu'en parallèle, la plus grosse entreprise du secteur qui pesait 50% du chiffre d'affaires en N-1 disparaît car elle est absorbée par une autre unité d'un secteur différent.

Une simple comparaison des résultats en niveau des deux années consécutives peut conduire à un diagnostic erroné. En effet, entre les deux années, le chiffre d'affaires du secteur a baissé de 45 %. On pourrait donc conclure au fait que le secteur se porte mal alors qu'en réalité les entreprises du secteur sont toutes en croissance. Deux messages contradictoires sont donc possibles.

A partir de ce constat et afin de contrebalancer les discours qui se basent sur les évolutions apparentes⁴, il a été décidé de prendre en compte, ou plutôt d'essayer de neutraliser les restructurations afin de rétablir des évolutions économiques sectorielles cohérentes entre 2 années successives. Pour ce faire, l'objectif est de recalculer la valeur des caractéristiques en N-1 en se plaçant dans la même structure en termes d'entreprise qu'en N. Il s'agit en quelque sorte de faire comme si la restructuration avait déjà eu lieu en N-1. Un nouveau concept d'unité statistique a donc été créé. Il s'agit de l'enveloppe de restructuration qui est géré à l'Insee dans un outil baptisé CITRUS⁵ [6].

Cette unité statistique se compose de deux listes d'entreprises ou Siren (entreprises de l'année N-1 avant la restructuration ; entreprises de l'année N après la restructuration).

⁴ Par évolution apparente il faut comprendre l'évolution entre les agrégats en niveau calculé en N-1 et N.

⁵ CITRUS (Coordination des Informations et des Traitements sur les Restructurations d'Unités Statistiques) est un système d'information sur les restructurations

Il est ensuite possible de calculer une APE pour l'enveloppe à partir des APE des unités légales qui la composent l'année N. Afin de neutraliser des changements de secteurs, l'APE de l'enveloppe en N-1 est la même qu'en N.

Les données de l'enveloppe sont calculées en cumulant les données des unités légales qui composent l'enveloppe en N et en N-1 et ce aussi bien pour les caractéristiques fiscales que pour les questionnaires d'enquêtes statistiques.

Pour les données N-1, on utilise un coefficient d'agrégation, permettant de prendre en compte les modifications des caractéristiques comptables, liées à la modification des flux économiques existant entre les unités concernées.

Ces données « fictives » vont ensuite être utilisées dans les calculs d'évolution en lieu et place des données des entreprises qui les composent.

2. Quelle décomposition et quelles sous populations ?

2.1. Les différents calculs d'évolution dans Esane

2.1.1. L'évolution apparente

Il s'agit d'un simple rapport des agrégats en niveau N / N-1.

La formule de cette évolution est la suivante :

$$\Delta_{N/N-1} = \frac{(A \text{ var}_N - A \text{ var}_{N-1})}{A \text{ var}_{N-1}} \quad \text{où } A \text{ var}_N \text{ est l'agrégat de la variable } A \text{ l'année } N \text{ (formule 1)}$$

2.1.2. L'évolution économique en comparable

L'objectif de l'évolution en comparable est de neutraliser les effets des restructurations en prenant en compte les enveloppes de restructuration venant de CITRUS.

Les entreprises prises en compte pour le calcul de l'agrégat N sur le secteur s en comparable sont :

- les entreprises actives du secteur s en N qui n'appartiennent pas à une enveloppe de restructuration N ;
- les entreprises actives qui appartiennent à une enveloppe de restructuration, enveloppe classée dans le secteur s en N.

Les entreprises prises en compte pour le calcul de l'agrégat N-1 sur secteur s en comparable sont :

- les entreprises actives du secteur s en N-1 qui n'appartiennent pas à une enveloppe de restructuration en N ;
- les entreprises actives en N-1 qui appartiennent à une enveloppe de restructuration, enveloppe classée dans le secteur s en N (l'information sur ces entreprises est consolidée sur le contour de l'enveloppe et est donc différente de la somme des entreprises pour les caractéristiques non additives. La somme obtenue sur les unités légales est pondérée par le coefficient d'agrégation).

2.2. Quels facteurs explicatifs ?

Comme le précise le document de travail « Les calculs d'évolution dans les enquêtes annuelles d'entreprise » [7], pour un secteur donné, il est possible de répartir les entreprises en 6 catégories et d'estimer le poids de chacune de ces catégories dans l'évolution économique du secteur.

Les différents facteurs d'évolution mis en évidence, et les catégories d'entreprises qui en découlent sont les suivants :

- La démographie d'entreprises (créations et cessations) :
on cherchera à mesurer par exemple l'impact de la disparition d'entreprises industrielles - extraction, textile...- ou de la création de start up, de l'immatriculation des panneaux solaires, la création du statut d'autoentrepreneurs dans l'évolution de leur secteur.
On cherche donc à isoler :
 - o Les entreprises naissantes (entreprises créées au cours de l'exercice N) : on s'intéressera aux entreprises créées ou réactivées, mais on ne prendra pas en compte les entrées de champ.
 - o Les entreprises cessées (entreprises cessées au cours de l'exercice N-1) : une entreprise cessée en cours d'année N sera considérée comme pérenne (même si elle n'a pas 12 mois d'activité pour l'année N), il en va de même pour les entreprises créées en N-1.
- Les restructurations d'entreprises: les entreprises⁶ en restructuration voient leur comportement économique modifié, ce qui impacte nécessairement les agrégats sectoriels. On retiendra donc pour notre analyse le contour de ces restructurations, appelé enveloppe.
- Les modifications de classement sectoriel :
On cherche ici à mesurer les entrées et sorties de champ : on entend ici par champ le champ d'observation sur lequel on se place. Si l'on considère un groupe donné de la NAF, on considérera les mouvements sectoriels de et vers le groupe comme des sorties et entrées de champ. Il s'agit donc ici d'identifier les entreprises ayant un classement différent en N-1 et N.
 - o Les entreprises en entrée de champ (entreprise active en N dans le secteur et active en N-1 mais dans un autre secteur, les deux secteurs étant compris dans le champ des enquêtes)
 - o Les entreprises en sortie de champ (entreprise active en N-1 dans le secteur et active en N mais dans un autre secteur, les deux secteurs étant compris dans le champ des enquêtes)
- Les pérennes⁷ : il s'agit des entreprises ayant connu une activité au cours des deux millésimes considérés, et n'ayant pas connu de modification de classement.

Le classement d'une unité donnée dans une de ces catégories dépend du niveau d'agrégation auquel on se place.

Une entreprise qui change d'APE au « niveau groupe » entre N et N-1, passant du 2830Z au 2841Z par exemple, sera considérée comme :

- pérenne éventuellement si on calcule des évolutions au niveau division de la NAF. Dans l'exemple, elle sera pérenne dans la division 28 ;
- comme en entrée de champ pour son groupe d'arrivée et en sortie de champ de son groupe de départ si on calcule des évolutions au niveau groupe de la NAF. Dans l'exemple, elle sera donc considérée en entrée du champ 284 et en sortie du champ 283.

Les deux dernières catégories reposent sur la pérennité ou la modification du classement sectoriel entre N-1 et N. Cela implique d'avoir réellement observé le classement sectoriel en N et en N-1. Ainsi, seules les données des entreprises ayant réellement répondu deux années successives pourront être

⁶ Une entreprise absorbée par une autre sera donc considérée comme faisant partie d'une restructuration bien qu'elle soit cessée en N. De la même façon une entreprise issue d'une scission totale sera considérée comme en restructuration et non naissante.

⁷ Le statut de pérenne est lié à l'activité de l'unité. Ainsi, une unité qui cesse au cours de l'exercice N mais qui a été active une partie de l'exercice sera considérée comme pérenne. Les durées d'exercice N et N-1 seront certes différentes mais l'unité est prise en compte dans les pérennes. Idem pour une unité créée en cours d'exercice N-1.

mobilisées pour nos estimations, puisque c'est uniquement pour ces deux populations que l'on dispose de cette information sur deux années consécutives⁸.

En effet, lorsqu'on observe au cours de l'année N une modification de classement, celle-ci peut être intervenue réellement beaucoup plus tôt, si l'entreprise n'a pas été interrogée depuis longtemps.

Initialement il avait été décidé de prendre en compte pour ces estimations les entreprises dont la réponse avait été extrapolée par N-1 (cette extrapolation étant jugée de bonne qualité pour le calcul des agrégats en niveau par exemple). Si cette décision avait l'avantage d'augmenter le nombre d'unités mobilisées pour nos calculs, elle avait l'inconvénient de biaiser les évolutions en les normalisant. L'extrapolation par N-1 empêche par essence les changements de secteur par exemple. C'est pourquoi il a été décidé in fine de se contenter des entreprises dont on dispose de la réponse les deux années pour nos estimations.

Comme précisé dans le paragraphe 1.2, le tirage de l'échantillon de l'EAP ne comprend pas de partie conservée, et ne permet donc pas de connaître les modifications sectorielles de l'année pour les entreprises appartenant au champ de l'industrie au lancement, alors que l'ESA permet de les connaître pour la partie conservée de l'échantillon.

Autrement dit, on est capable d'estimer les mouvements sectoriels des entreprises issues des secteurs du commerce, des services... y compris vers les secteurs industriels, mais pas les mouvements sectoriels inverses.

Le poids important des entreprises de la partie exhaustive limite le biais pour les agrégats économiques, en revanche, les estimations de nombre d'entreprises pourront être biaisées de façon plus importante.

On limitera donc l'analyse aux secteurs couverts par l'ESA, avec une possible sous-estimation, puisque les éventuelles entreprises venant de l'industrie ne seront pas prises en compte.

⁸ Cette décision nous pousse donc à exclure de notre analyse les unités entrantes ou sortantes du champ de l'enquête. Pour ces unités en effet nous ne disposons au mieux que d'une information sur une des deux années.

2.3. La décomposition de l'évolution

La décomposition de l'évolution économique peut donc s'écrire de la façon suivante :

$$A_N = P_N + R_N + N_N + EC_N$$

$$\left\{ \begin{array}{l} P_N = \text{agrégat des entreprises pérennes} \\ R_N = \text{agrégat des entreprises en restructuration} \\ N_N = \text{agrégat des naissantes} \\ EC_N = \text{agrégat des entrées de champ en } n \end{array} \right. \quad \text{(formule 2)}$$

$$A_{N-1} = P_{N-1} + R_{N-1} + M_{N-1} + SC_N$$

$$\left\{ \begin{array}{l} P_{N-1} = \text{agrégat } N-1 \text{ des entreprises pérennes} \\ R_{N-1} = \text{agrégat } N-1 \text{ des entreprises en restructuration} \\ M_{N-1} = \text{agrégat des cessées} \\ SC_N = \text{agrégat des sorties de champ en } n \end{array} \right.$$

$$\Delta_{N/N-1} = \frac{(A_N - A_{N-1})}{A_{N-1}} = \frac{(P_N - P_{N-1})}{A_{N-1}} + \frac{(R_N - R_{N-1})}{A_{N-1}} + \frac{(N_N - M_{N-1})}{A_{N-1}} + \frac{(EC_N - SC_N)}{A_{N-1}} \quad \text{(formule 3)}$$

L'estimation de cette décomposition est toutefois complexe car il est difficile de définir la notion de « pérennes » lorsque l'on utilise les résultats d'un échantillon pour calculer des agrégats.

La notion de pérennité, d'entrée et de sortie de champ ne peut être calculée que sur :

- la partie exhaustive de l'échantillon
- la partie conservée de l'échantillon
- les entreprises de la partie renouvelée ayant été tirées les deux années.

Comme précisé ci-dessus, il est nécessaire de disposer de deux années d'observation pour pouvoir caractériser ces unités. Dans le document de travail « Les calculs d'évolution dans les enquêtes annuelles d'entreprise » [7], il est proposé de prendre également en compte les entreprises appartenant à la partie renouvelée de l'échantillon, mais qui seraient tout de même interrogées les deux années. Ceci est possible en statistique d'entreprises car certaines strates contiennent peu d'unités et parce qu'il n'y a pas de coordination négative pour la partie renouvelée de l'échantillon.

La notion de « naissante » peut être estimée sur :

- la partie exhaustive de l'échantillon
- l'échantillon des naissantes tiré en N.

La notion de cessée peut être estimée sur :

- La totalité de l'échantillon N-1.

La notion de restructuration peut être estimée :

- Sur la totalité de l'échantillon N.

On souhaite utiliser des informations réellement fournies par les entreprises, et non extrapolées.

On ne retiendra donc uniquement les entreprises ayant réellement répondu deux années de suite pour les pérennes, et ayant répondu pour l'année concernée pour les cessées (année N-1) et les créées (année N).

Cependant, dans les populations des créées et des cessées, on observe un plus grand taux de non réponse, y compris parmi les entreprises appartenant à la strate exhaustive. Les entreprises non répondantes appartenant à la strate exhaustive font l'objet d'une extrapolation individuelle. Il a été décidé de les conserver afin d'utiliser un nombre suffisant d'entreprises (en particulier pour les cessées).

Au vu de ces différentes contraintes il a été décidé de conserver les entreprises suivantes :

- Les répondantes pour les deux années de la partie exhaustive de l'échantillon ;
- Les répondantes pour les deux années de la partie conservée de l'échantillon (pour la partie non exhaustive) ainsi que celles appartenant à la partie renouvelée, mais tirées dans les deux échantillons ;
- Les entreprises naissantes répondantes ou les entreprises extrapolées appartenant à la strate exhaustive de l'échantillon des naissantes ;
- Les entreprises cessées en cours d'année N-1 répondantes de l'échantillon N-1, ou les entreprises extrapolées appartenant à la strate exhaustive.

Afin d'obtenir une estimation de l'ensemble du champ, les agrégats obtenus à partir des populations décrites ci-dessus seront pondérés afin de représenter l'ensemble du champ.

L'agrégat d'une caractéristique X pour le secteur s peut donc être calculé de la façon suivante :

$$Agr_X_s^N = Agr_X_exh_{s_rep}^N + Agr_X_exh_{s_non_rep}^N + Agr_X_cons_s^N + Agr_X_nais_s^N + Agr_X_ren_{sn \cap sn-1}^N + Agr_X_ren_{sn}^N \quad \text{(formule 4)}$$

où

$$\left\{ \begin{array}{l} Agr_X_s^N = \text{agrégat de la caractéristique X sur le secteur s} \\ Agr_X_exh_{s_rep}^N = \text{agrégat de la caractéristique X calculé sur la partie exhaustive de l'échantillon du secteur s (répondantes)} \\ Agr_X_exh_{s_non_rep}^N = \text{agrégat de la caractéristique X calculé sur la partie exhaustive de l'échantillon du secteur s (non répondantes extrapolées individuellement)} \\ Agr_X_cons_s^N = \text{agrégat de la caractéristique X calculé sur la partie renouvelée de l'échantillon du secteur s} \\ Agr_X_nais_s^N = \text{agrégat de la caractéristique X calculé sur la partie des naissantes de l'échantillon du secteur s} \\ Agr_X_ren_{sn \text{ inter } sn-1}^N = \text{agrégat de la caractéristique X calculé sur la partie renouvelée de l'échantillon du secteur s, pour les entreprises également tirées en N-1 et répondantes les deux années} \\ Agr_X_ren_{sn}^N = \text{agrégat de la caractéristique X calculé sur le reste de la partie renouvelée de l'échantillon du secteur s} \end{array} \right.$$

On notera par la suite $Agr_X_cons_large_s^N$ l'agrégat de la partie conservée large, i. e.

$$Agr_X_cons_s^N + Agr_X_ren_{sn \cap sn-1}^N$$

Par un simple jeu de réécriture de la formule 4 ci-dessus, on peut écrire l'agrégat N d'une caractéristique X pour le secteur s de la façon suivante :

$$Agr_X_s^N = Agr_X_exh_{s_rep}^N * \frac{Agr_X_exh_{s_rep}^N + Agr_X_exh_{s_non_rep}^N}{Agr_X_exh_{s_rep}^N} + Agr_X_nais_s^N + \frac{(Agr_X_cons_large_s^N + Agr_X_ren_s^N)}{Agr_X_cons_large_s^N} * Agr_X_cons_large_s^N$$

(formule 5)

$$\text{On pose ensuite } K_s^N ech = \frac{Agr_X_cons_large_s^N + Agr_X_ren_s^N}{Agr_X_cons_large_s^N} \quad \text{(formule 6)}$$

$$\text{et } K_s^N exh = \frac{Agr_X_exh_{s_rep}^N + Agr_X_exh_{s_non_rep}^N}{Agr_X_exh_{s_rep}^N} \quad \text{(formule 6 bis)}$$

Autrement dit les K_s^N représentent l'inverse du poids de la partie des doublement répondantes dans les parties exhaustives, et conservée large de l'échantillon dans l'agrégat de l'ensemble de ces parties.

En décomposant chaque agrégat défini dans la formule 4 on obtient les formules détaillées ci-dessous

$$Agr_X_exh_{s_rep}^N = \sum_{i \in exh} X_i^N * (\mathbb{1}_{i \in perenne_s} + \mathbb{1}_{i \in naissante_s} + \mathbb{1}_{i \in entchamp_s} + \mathbb{1}_{i \in restruct_s}) \quad \text{(formule 7)}$$

où

$$\left\{ \begin{array}{l} X_i^N \text{ est la valeur de l'année N de la caractéristique X de l'unité i} \\ \mathbb{1}_{i \in perenne_s} = \begin{cases} 1 \text{ si l'unité i n'est pas en restructuration et est pérenne dans le secteur s} \\ 0 \text{ sin on} \end{cases} \\ \mathbb{1}_{i \in naissante_s} = \begin{cases} 1 \text{ si l'unité i n'est pas en restructuration et est naissante dans le secteur s} \\ 0 \text{ sin on} \end{cases} \\ \mathbb{1}_{i \in entchamp_s} = \begin{cases} 1 \text{ si l'unité i n'est pas en restructuration et est dans le secteur s en N} \\ \text{et dans un autre en N-1} \\ 0 \text{ sin on} \end{cases} \\ \mathbb{1}_{i \in restruct_s} = \begin{cases} 1 \text{ si l'unité i appartient à une enveloppe de restructuration de l'année N du secteur s} \\ 0 \text{ sin on} \end{cases} \end{array} \right.$$

$$Agr_X_cons_large_s^N = \sum_{i \in cons_large} w_i^N * X_i^N * (\mathbb{1}_{i \in perenne_s} + \mathbb{1}_{i \in entchamp_s}) \quad \text{(formule 8)} \text{ où } w_i^N \text{ est le poids après calage de l'unité i pour l'année N.}$$

NB. : Par définition il n'y a ni naissante ni entreprise en restructuration dans la partie conservée. Les unités faisant partie d'une restructuration sont toujours « basculées » dans la partie exhaustive.

$$Agr_X_nais_s^N = \sum_{i \in nais} w_i^N * X_i^N * (\mathbb{1}_{i \in naissante_s}) \quad \text{(formule 9)}$$

$$Agr_X_ren_s^N = \sum_{i \in ren} w_i^N * X_i^N * (\mathbb{1}_{i \in ren}) \quad \text{(formule 10)}$$

où

$$\mathbb{1}_{i \in ren} = \begin{cases} 1 \text{ si l'unité i n'est pas en restructuration et est dans l'échantillon renouvelé du secteur s} \\ 0 \text{ sin on} \end{cases}$$

2.4. Les différentes sous populations à isoler

On peut donc réécrire de la façon suivante les différentes sous populations à mettre en évidence.

2.4.1. Les entreprises pérennes

Ainsi, la valeur de la caractéristique en N des entreprises pérennes sera estimée de la façon suivante (utilisation des formules 5, 6, 7 et 8).

$$P_X_s^N = K_s^N exh(1) * \sum_{i \in exh} X_i^N * \mathbb{1}_{i \in perenne_s} + K_s^N ech(1) * \sum_{i \in cons_large} w_i^N * X_i^N * \mathbb{1}_{i \in perenne_s} \quad \text{(formule 11)}$$

2.4.2. Les entreprises en restructuration

$$R_X_s^N = \sum_{i \in \text{exh}} X_i^N * \mathbb{I}_{i \in \text{restruct}_s} \quad (\text{formule 12})$$

2.4.3. Les entreprises créées et cessées

$$N_X_s^N = \sum_{i \in \text{exh}} X_i^N * \mathbb{I}_{i \in \text{naissante}_s} + \sum_{i \in \text{nais}} w_i^N * X_i^N * \mathbb{I}_{i \in \text{naissante}} \quad (\text{formule 13})$$

$$C_X_s^{N-1} = \sum_{i \in \text{exh}} X_i^{N-1} * \mathbb{I}_{i \in \text{cessées}_s} + \sum_{i \in \text{ech}} w_i^{N-1} * X_i^{N-1} * \mathbb{I}_{i \in \text{cessées}_s} \quad (\text{formule 14})$$

2.4.4. Les entreprises en entrée et sortie de champ

$$EC_X_s^N = K_s^N \text{exh}(1) * \sum_{i \in \text{exh}} X_i^N * \mathbb{I}_{i \in \text{entchamp}_s} + K_s^N \text{ech}(1) * \sum_{i \in \text{cons_l arg e}} w_i^N * X_i^N * \mathbb{I}_{i \in \text{entchamp}_s} \quad (\text{formule 15})$$

$$SC_X_s^{N-1} = K_s^N \text{exh}(1) * \sum_{i \in \text{exh}} X_i^{N-1} * \mathbb{I}_{i \in \text{sortiechamp}_s} + K_s^{N-1} (\text{ech}) * \sum_{i \in \text{cons_l arg e}} w_i^{N-1} * X_i^{N-1} * \mathbb{I}_{i \in \text{sortiechamp}_s} \quad (\text{formule 16})$$

2.4.5. Evolution, poids et contribution

A partir de la formule 3, et des formules ci dessus, on pourra calculer :

- le poids d'une catégorie en niveau : le poids des pérennes en N-1 $\frac{P_{N-1}}{A_{N-1}}$

- l'évolution de certaines catégories (les pérennes et les restructurations) :

$$\Delta P_{N/N-1} = \frac{(P_N - P_{N-1})}{P_{N-1}} \quad (\text{formule 17})$$

- la « contribution » de chacun des éléments : $CP_{N/N-1} = \frac{(P_N - P_{N-1})}{A_{N-1}}$ (formule 18)

- ou encore le poids dans l'évolution : $\frac{(P_N - P_{N-1})}{(A_N - A_{N-1})}$

3. Quel poids pour les unités ?

Le traitement de la non réponse étant fait par ré pondération, les poids finaux peuvent être différents pour une unité entre N et N-1, y compris pour la partie conservée de l'échantillon où le poids initial était le même les deux années.

Le poids final est influencé par le comportement de réponse des entreprises, qui peut être différent d'une année sur l'autre.

L'entreprise elle même peut être répondante une année mais pas l'autre.

Tableau 3 : Schéma de réponse pour deux millésimes successifs

Entreprises appartenant à la partie conservée de l'échantillon	Poids au lancement en N-1 et en N	Répondantes en N-1	Poids final en N-1	Répondantes en N	Poids final en N
◆	10	◆	16,6	◆	25
◆	10		0		0
◆	10		0		0
◆	10		0	◆	25
◆	10	◆	16,6	◆	25
◆	10	◆	16,6		0
◆	10	◆	16,6		0
◆	10		0	◆	25
◆	10	◆	16,6		0
◆	10	◆	16,6		0

Pour les entreprises de la partie conservée, le rapport des poids finaux N et N-1 varie de 0,1 à 4.

Tableau 4 : Distribution des rapports de poids finaux N et N-1 des entreprises de la partie conservée

	Minimum	1 ^{er} décile	1er quartile	Médiane	3ème quartile	9 ^{ème} décile	Maximum
Rapport des poids 2010/2009	0,079	0,799	0,897	0,991	1,105	1,231	4,608

NB : les entreprises non répondantes pour un des deux millésimes ont été exclues. Ce tableau ne porte donc que sur les entreprises ayant un poids strictement positif et différent de 1 pour les deux millésimes.

Tableau 5 : Répartition des entreprises selon leur réponse sur deux millésimes successifs

	En part du nombre d'entreprises				
	Réponse en N-1 seulement	Réponse en N seulement	Réponse en N et en N-1	Aucune réponse	Ensemble
Partie conservée 2008/2009	14,7%	7,2%	55,1%	23,0%	100 %
Partie conservée 2009/2010	11,4%	10,2%	53,1%	25,3%	100 %

Dans ce cas, quel poids doit on conserver pour l'entreprise, et comment gérer les différents cas qui se présentent ?

3.1. Option 1 : conserver le poids N pour les unités en N mais également en N-1.

Sur la partie conservée les poids au lancement sont les mêmes pour les deux années. Cependant le traitement de la non réponse étant fait par repondération, les poids finaux ne sont pas nécessairement égaux !

Puisque l'on conserve uniquement des entreprises ayant réellement répondu deux années consécutives, on pourrait appliquer alors aux entreprises en N-1 le poids final N, et de même le coefficient K_s serait alors calculé de façon identique pour les deux années. L'objectif, en agissant ainsi, est de prendre en compte les « vraies » évolutions observées sur les entreprises interrogées pour deux exercices successifs.

Cependant, suite aux traitements sur les poids différents pour les deux années, on perdrait le calage pour l'année N-1. Ce choix conduirait à s'éloigner plus fortement de l'agrégat connu pour l'année N-1.

On observe aussi des basculements de la partie échantillonnée vers la partie exhaustive et inversement, l'utilisation d'un poids N « tord » ainsi la représentation en N-1.

3.2. Option 2 : conserver pour chacun des millésimes les poids du millésime

L'autre option qui se présente est de conserver pour chaque unité le poids tel qu'il a été déterminé lors des traitements de fin d'enquête.

On conserve dans ce cas le calage effectué pour chacun des millésimes, on est plus proche des agrégats en niveau.

L'option 2 « poids millésimés » repose sur la confiance que l'on fait à l'échantillonnage. Elle fait l'hypothèse que les mécanismes de réponse sont ignorables, ou du moins, s'ils sont non ignorables, qu'ils le sont de la même façon chaque année. Autrement dit, puisque l'on cherche à mesurer une évolution, que la conjoncture économique n'implique pas un mécanisme de réponse différent selon les catégories que l'on cherche à mesurer.

L'option 1 « poids N » privilégie une approche « cylindrique » de l'évolution, mais nous oblige à casser le calage des poids, et à nous « déconnecter » des agrégats publiés. Cette option serait sans doute préférable avec un échantillon de taille plus importante.

3.3. Comment représenter l'ensemble de l'échantillon ?

L'utilisation d'un demi échantillon nous oblige à corriger les agrégats obtenus par le biais d'un coefficient K_s (cf. formule 6 ci-dessus).

Selon l'option retenue pour le poids individuel de l'entreprise, le calcul du coefficient K_s sera différent. Ainsi, si l'on choisit d'utiliser les poids N pour les deux millésimes, on calculera K_s à partir des données N pour les deux années.

Si l'on choisit de conserver les poids issus des traitements pour chacune des deux années, alors les K_s utilisés seront relatifs aux données du millésime concerné.

On aurait pu choisir également un K_s indépendant de la variable en prenant en compte seulement les poids, ou prendre arbitrairement un K_s égal à 2. Ces éventualités n'ont pas été testées.

A priori, on souhaitait conserver l'option du poids N pour les deux millésimes.

Cependant, le choix des poids et du K_s correspondant a une influence sur les résultats.

3.4. L'influence du choix de poids sur les résultats

Les agrégats utilisés pour décomposer l'évolution (formules 4 et 5) utilisent uniquement les données de l'enquête, alors que les agrégats publiés par ailleurs sont calculés selon une méthodologie composite [1]. De plus, les agrégats publiés le sont souvent sur un champ un peu plus large que celui des enquêtes. Il en résulte parfois un léger écart, entre l'agrégat publié, et l'agrégat utilisé pour décomposer l'évolution.

En outre, si on choisit d'utiliser des poids N pour les deux millésimes, et des K_s calculés sur l'année N pour les deux millésimes, on s'écarte d'autant plus de l'agrégat N-1 publié.

Le ratio de l'agrégat N se comporte de façon similaire (mais est indifférent à l'option retenue pour les poids).

En conséquence, le choix de l'option « poids millésimé » améliore l'agrégat calculé en N-1 pour la décomposition, et améliore donc l'écart entre l'évolution publiée et l'évolution que l'on sait décomposer.

Ainsi pour des raisons de pertinence des résultats, il a été décidé de conserver les poids de chacun des millésimes pour l'estimation de la décomposition des évolutions.

4. Résultats

Dans cette partie, nous présenterons quelques résultats, en particulier en essayant de valider de façon empirique les choix effectués, et en montrant leurs limites. Les travaux ont été menés sur deux variables (chiffre d'affaires, valeur ajoutée) et au niveau groupe (3 premiers chiffres) de la NAF (une centaine de groupes après exclusion des secteurs couverts par l'EAP).

4.1. Ecarts entre évolution apparente et évolution décomposée

Comme précisé dans le point 3.4, certaines évolutions décomposées s'éloignent parfois de façon notable de l'évolution publiée par ailleurs. Les raisons principales étant dues au fait que les évolutions publiées concernent un champ plus large d'une part, et d'autre part que les évolutions estimées pour la décomposition ne prennent en compte qu'une partie de l'échantillon.

Toutefois, l'idée de cette décomposition est bien d'explicitier une tendance et d'en fournir les facteurs explicatifs, et non de calculer au demi point près le poids de l'une ou l'autre des catégories.

Toutefois, afin de rester un minimum « crédible » vis à vis des utilisateurs, il a été décidé de se limiter à l'analyse des décompositions des secteurs vérifiant les conditions suivantes :

- l'évolution décomposée doit être de même signe que l'évolution calculée d'après les agrégats publiés, ou l'écart entre les évolutions doit être inférieur à 1 point ;
- l'écart entre ces évolutions ne doit pas être trop important (5 points par exemple) ;
- le rapport des deux évolutions doit être compris entre 1/2 et 3/2.

Selon ces conditions, on pourrait utiliser l'information pour 70 % des groupes. Pour 40 % des groupes l'écart entre les évolutions est inférieur à 1 point, pour 20 % il est inférieur à 2 points et pour 24 %, l'écart est compris entre 2 et 5 points.

Une réflexion est encore en cours afin de trouver une solution pour les groupes non exploitables selon ces conditions.

4.2. Nombre d'unités participant aux différents éléments des évolutions

Selon les secteurs, l'analyse finale de l'évolution peut porter sur assez peu d'unités.

Ainsi, l'analyse de la décomposition des évolutions repose sur 52 000 entreprises, dont 15 500 issues de l'échantillon . Si cela peut paraître beaucoup, cette situation cache des inégalités sectorielles et peut donc poser des problèmes de fiabilité des résultats. Ainsi, il y a moins de 10 entreprises de l'échantillon (partie sondée) qui participent au calcul (pour 20 groupes) et pour 32 d'entre eux, il y a moins de 30 entreprises.

Il est cependant difficile de fixer un seuil raisonnable en dessous duquel on n'accorderait plus de crédit à l'analyse, car le poids de l'exhaustif d'une part, et le poids moyen des entreprises interrogées d'autre part varient fortement d'un secteur à l'autre. Ainsi les secteurs des transports en commun ont très peu d'unités, mais qui sont toutes interrogées exhaustivement.

On veillera toutefois à avertir l'utilisateur lorsque les populations utilisées sont de très petite taille.

4.3. Présentation d'un résultat

De façon synthétique, on pourrait présenter le résultat de la façon suivante :

Tableau 6 : Décomposition de l'évolution 2010/2009 du CA et de la VA pour le secteur « Intermédiaires du commerce de gros » (461)

Variable	Agrégat 2010 (en keuros)	Evolution en comparable	Evolution décomposée	Evolution expliquée par (éléments de la formule 18)				Evolution de la population des pérennes (formule 17)
				Pérennes	Restructurations	Créations Cessations	Entrées et sorties de champ	
CA	122 112 711	4,13 %	4,06 %	6,44 %	- 0,02 %	- 0,30 %	- 2,05 %	7,28 %
Va	10 436 913	- 5,43 %	- 4,60%	2,65 %	- 0,03 %	- 0,06 %	- 7,29 %	3,38 %

Source : ESANE 2009 et 2010

Dans l'exemple ci-dessus, les évolutions « vraie » et décomposée sont relativement proches. L'évolution du chiffre d'affaires est expliquée principalement par les pérennes, alors que le secteur connaît quelques sorties de champ.

En revanche, les mouvements sectoriels expliquent la baisse de la valeur ajoutée du secteur, alors même que la valeur ajoutée des entreprises pérennes augmente.

Si l'on analyse de façon plus générale les résultats (pour les groupes retenus après les limites énoncées dans les points précédents), on observe :

- pour deux secteurs sur trois, les pérennes expliquent la moitié ou plus de l'évolution ;
- on retrouve un peu plus d'une vingtaine de secteurs pour lesquelles les pérennes ont un comportement opposé à celui du secteur. Cela arrive lorsque l'évolution est proche de 0, et dans ce cas l'évolution des pérennes est finalement assez proche de l'évolution du secteur, bien que de signe opposé. Dans les autres cas, il s'agit souvent de secteurs où les pérennes voient leur activité diminuer, l'évolution sectorielle étant alors expliquée par les restructurations ;
- le poids de la démographie d'entreprises est souvent imputable en très grande partie à une seule des catégories (entreprises cessées en N-1 versus entreprises créées en N) ;
- en revanche, l'évolution due aux entrées et sorties de champ est souvent partagée entre les deux phénomènes (sorties versus entrées).

Conclusion

Bien que partant d'un problème facilement énonçable (comment expliquer l'évolution d'un secteur selon certaines catégories de populations), la mise en œuvre est assez complexe, en raison des difficultés pour identifier les différentes catégories et leurs poids, difficultés liées à l'échantillonnage et au traitement de la non réponse.

Ces travaux nous permettent toutefois d'identifier quelles sont les principales tendances expliquant les évolutions sectorielles. Ces messages sont extrêmement utiles pour définir le véritable état de santé des différentes activités.

Dans le climat de crise actuel, ils sont d'autant plus importants car ils sont autant d'éléments pouvant permettre d'orienter des politiques économiques.

De plus ces indicateurs seront également un outil très utile au moment de la vérifications des résultats avant leur publication. En effet, une évolution très atypique des entreprises pérennes par exemple pourrait être le fait d'une erreur de données. Cette évolution atypique méritera en tout cas d'être examinée avant toute publication pour en comprendre la cause.

Les choix méthodologiques effectués sont certes discutables mais on en connaît a priori les limites. Un certain nombre de ces choix pourraient être approfondis, en particulier le choix du poids que l'on souhaite accorder à chacune d'entre elles (repondération individuelle par un coefficient de type inverse du poids de la partie conservée dans l'échantillon calculé au niveau des strates de réponse, ajustement de l'agrégat dans un second temps). Le calcul de poids longitudinaux n'a pas été étudié pour l'instant.

On pourrait également envisager le calcul d'évolution purement cylindrées que l'on appliquerait ensuite (de façon absolue aux vrais niveaux, ou en structure à la vraie évolution).

Enfin, on pourrait s'intéresser à la publication d'indicateurs de tendance, en identifiant pour un secteur la principale source d'évolution et son poids dans l'évolution totale, ou encore au calcul d'un score de confiance selon les différents éléments statistiques présentés plus haut (valeur des K_s , ratio des agrégats, ratio des évolutions, nombre d'unités participant au calcul, poids moyen des entreprises échantillonnées...).

Bibliographie

[1] Esane, le dispositif rénové de production des statistiques structurelles d'entreprises Philippe Brion, Courrier des statistiques n°130, 02/05/2011

[2] L'utilisation combinée de données d'enquête et de données administratives pour la production des statistiques structurelles d'entreprises, Brion Ph., papier présenté aux JMS de l'Insee 2009

[3] Esane ou les malheurs de l'estimation composite : comment gérer les valeurs négatives d'estimateurs par différence ?, Gros E. papier présenté aux JMS de l'Insee 2012

[4] Winsorisation dans les enquêtes annuelles auprès des entreprises françaises, Brion Ph., Guggemos F., papier présenté au Sixième Colloque francophone sur les Sondages , 2010

[5] Esane : À la recherche d'une cohérence maximale des données multi-sources sur les entreprises par le biais de micro et macro contrôles, Haag O., papier présenté aux JMS de l'Insee 2012

[6] CITRUS - Système d'information sur les restructurations d'entreprises , M. Beauvois, Courrier des statistiques n°95-96, 2000

[7] Les calculs d'évolution dans les enquêtes annuelles d'entreprise, Rivière P., Document de travail Insee n°E9801