
***DEUXIÈME PARTIE :
RÉALISATION DE L'ENQUÊTE***

DEUXIÈME PARTIE : RÉALISATION DE L'ENQUÊTE

La réalisation d'une enquête comprend différentes phases faisant intervenir l'équipe responsable de l'enquête au niveau national (située au Département de la Conjoncture de l'Insee) et une équipe d'informaticiens (basée dans plusieurs Centres Nationaux Informatiques de l'Institut).

Les différentes étapes de la réalisation de l'enquête *stricto sensu* (hors diffusion, interprétation et utilisation des résultats) sont les suivantes :

- la constitution et le renouvellement de l'échantillon ;
- la collecte et la saisie des données ;
- la constitution des données individuelles et leur contrôle ;
- le dépouillement de l'enquête ;
- la correction des variations saisonnières ;
- le stockage des résultats agrégés.

2.1. - La constitution et le renouvellement de l'échantillon

La constitution initiale de l'échantillon d'artisans à interroger résulte d'un sondage stratifié sans remise à partir de l'Enquête Annuelle d'Entreprise (EAE) la plus récente au moment du lancement de l'enquête (prochainement l'Enquête Sectorielle Annuelle ESA). Cet échantillon est pour l'essentiel suivi d'une enquête à l'autre (panel d'entreprises). Toutefois, en raison des évolutions économiques (disparitions ou restructurations d'entreprises) et de la nécessité de conserver un échantillon de taille adéquate, l'échantillon est régulièrement actualisé.

Ainsi, l'échantillon de l'enquête de conjoncture dans l'artisanat du bâtiment est renouvelé partiellement une fois par an, selon des modalités identiques d'une année à l'autre. La base de sondage est la dernière EAE connue (prochainement ESA). Lors de ce renouvellement, on détermine un échantillon théorique de 1 500 entreprises, selon un sondage stratifié suivant un croisement « taille de l'entreprise » en termes d'effectifs salariés et « secteur d'activité ». Le nombre d'entreprises de chaque strate de l'échantillon est proportionnel au poids en termes de chiffres d'affaires de la strate sur l'ensemble du secteur de l'artisanat du bâtiment.

Dans un second temps, on détermine le nombre d'entreprises à tirer pour compléter l'échantillon de l'année précédente une fois ce dernier toiletté (c'est-à-dire après suppression des entreprises disparues, restructurées, mauvaises répondantes, etc.). Les nouvelles entreprises sont choisies par tirage aléatoire simple sans remise dans chaque strate.

Entre deux campagnes de renouvellement, l'Insee effectue en outre un travail permanent de maintien de l'échantillon (mises à jour en continu des adresses des entreprises enquêtées et des correspondants au sein de ces dernières, suivi des restructurations et des cessions d'entreprises, etc.).

Davantage d'éléments sur l'échantillon sont donnés dans la fiche technique récapitulative de l'enquête, en annexe 1.

2.2. - La collecte et la saisie de l'information

2.2.1. - La collecte des données

Les questionnaires relatifs à l'enquête du mois m sont envoyés aux entreprises constituant l'échantillon vers le 27 du mois $m-1$. Les entreprises adressent en retour leur réponse à l'Insee, où est effectuée la saisie des questionnaires. Les réponses doivent y parvenir vers le 7 du mois de l'enquête.

À chaque enquête, un questionnaire de rappel est envoyé aux entreprises n'ayant pas répondu à la date de retour indiquée sur le questionnaire initial.

Depuis janvier 2008, un nouveau mode de collecte a été mis en place pour les entreprises, en parallèle de la collecte classique par voie postale. Il s'agit d'un questionnaire électronique accessible depuis le site internet : <http://conjoncture.entreprises.insee.fr>. Les entreprises ont le choix du mode de réponse : par papier ou par internet.

2.2.2. - La saisie des données individuelles et leur contrôle

La saisie des données est effectuée à l'Insee (Division des Enquêtes de Conjoncture), au fur et à mesure de l'arrivée des questionnaires.

Plusieurs types de contrôles des données sont opérés au sein de cette Division :

- contrôles sur la cohérence des données structurelles : ces contrôles portent essentiellement sur la cohérence des chiffres d'affaires globaux et des chiffres d'affaires par type de travaux, sur les effectifs salariés ainsi que sur les évolutions de ces données structurelles d'une année sur l'autre ;
- contrôles sur les codages autorisés lors de la saisie des réponses conjoncturelles, des valeurs précises étant autorisées selon les réponses des entreprises ;
- contrôles sur les nomenclatures : une table de nomenclature de saisie a été introduite dans l'application de saisie. Il s'agit de la table de nomenclature d'activités.

2.2.3. - La collecte par internet

Depuis janvier 2008, les artisans ont la possibilité de répondre à l'enquête de conjoncture dans l'artisanat du bâtiment par internet. Il leur suffit de se connecter sur le site internet et de s'identifier à l'aide de leurs code d'accès et mot de passe fournis par l'Insee. Ils répondent ensuite au questionnaire de manière totalement similaire à la réponse par papier.

Le site internet inclut un certain nombre d'explications et de services facilitant les réponses des entreprises, ainsi qu'un retour d'informations enrichi - Cf. alinéa 3.2.2. *infra*. Ce nouveau mode de collecte est déjà quasiment généralisé à l'ensemble des enquêtes de conjoncture.

Lorsque les entreprises répondent par internet, leurs réponses sont contrôlées. Les questionnaires bien remplis au regard de ces contrôles sont déversés directement dans les fichiers de données individuelles. Les questionnaires présentant des anomalies font l'objet d'une intervention manuelle.

2.3. - L'organisation des données individuelles

Les données individuelles sont stockées dans une base de données commune à toutes les enquêtes de conjoncture (Industrie, Bâtiment, Services, ...). Cette base de données contient :

- des informations décrivant les entreprises : numéro Siren, raison sociale, adresse, ... ;
- les données structurelles de chaque entreprise : chiffre d'affaires, effectifs salariés, code Naf de l'entreprise, chiffre d'affaires par type de travaux ;
- les données conjoncturelles : réponses aux questions qualitatives.

2.4. - Le dépouillement

Le dépouillement d'une question conjoncturelle correspond au calcul des niveaux agrégés à partir des données individuelles. Il s'effectue en deux étapes. On procède d'abord à l'agrégation des réponses au niveau de chaque strate élémentaire (cf. *infra*, alinéa 2.4.1). Dans un second temps, on agrège les résultats ainsi obtenus en tenant compte de la structure du champ de l'enquête.

La première étape du dépouillement (agrégation primaire) utilise uniquement des données de l'enquête : réponses aux questions conjoncturelles et données structurelles, qui jouent le rôle de coefficients de pondération. La seconde étape du dépouillement fait appel à des données externes (données de l'EAE, et prochainement de l'ESA), qui jouent le rôle de coefficients de redressement.

Toutes les questions d'une enquête sont dépouillées deux fois : une première fois lors de la publication de l'enquête courante et une seconde fois au moment de la publication de l'enquête suivante. Ce procédé est motivé par la prise en compte des réponses d'entreprises retardataires.

2.4.1. - La stratification

Une strate élémentaire est définie par un secteur d'activité qui regroupe plusieurs postes de la Naf 700, croisé avec une tranche de taille¹⁶. On distingue 4 secteurs différents (cf. *infra*, alinéa 1.4) et 3 tranches de taille : il existe donc $4 \times 3 = 12$ strates élémentaires.

2.4.2. - Le dépouillement des questions et le calcul des soldes d'opinion

2.4.2.1. - Agrégation primaire des résultats

L'agrégation primaire consiste à calculer des moyennes pondérées des réponses individuelles au sein de chaque strate élémentaire. Le coefficient de pondération est une donnée fournie par l'entreprise. Ce coefficient est fondé :

- sur l'effectif employé pour les questions portant sur les effectifs passés et prévus ainsi que pour la question sur les difficultés de recrutement ;
- sur le chiffre d'affaires (total, par type de travaux ou par type de clientèle) pour les autres questions.

¹⁶ Rappel de la définition des tranches de taille, spécifique à l'enquête : - tranche 0 : 0 salarié - tranche 1 : 1 à 5 salariés - tranche 2 : 6 à 10 salariés.

2.4.2.1.1 - Questions trimodales

Il s'agit de questions qualitatives, appelant une réponse positive (en hausse, meilleure, etc.), intermédiaire (stable, normal, etc.) ou négative (en baisse, plus mauvaise, etc.)¹⁷.

Sont calculés les pourcentages pondérés de réponses « en hausse », « stable » et « en baisse » pour chaque question et chaque strate élémentaire¹⁸ :

$$\% HAUSSE = 100 \times \frac{NP \text{ « en hausse »}}{NP} \quad (1)$$

$$\% STABLE = 100 \times \frac{NP \text{ « stable »}}{NP} \quad (2)$$

$$\% BAISSSE = 100 \times \frac{NP \text{ « en baisse »}}{NP} \quad (3)$$

avec :

$NP \text{ « en hausse »}$ = Somme des pondérations des réponses « en hausse »,

$NP \text{ « stable »}$ = Somme des pondérations des réponses « stable »,

$NP \text{ « en baisse »}$ = Somme des pondérations des réponses « en baisse »,

et $NP = NP \text{ « en hausse »} + NP \text{ « stable »} + NP \text{ « en baisse »}$.

Pour les perspectives générales d'activité, on calcule également les pourcentages non pondérés de réponses « en hausse », « stable » et « en baisse » pour chaque strate élémentaire :

$$\% HAUSSE_{\text{non pondéré}} = 100 \times \frac{N \text{ « en hausse »}}{N} \quad (4)$$

$$\% STABLE_{\text{non pondéré}} = 100 \times \frac{N \text{ « stable »}}{N} \quad (5)$$

$$\% BAISSSE_{\text{non pondéré}} = 100 \times \frac{N \text{ « en baisse »}}{N} \quad (6)$$

avec :

$N \text{ « en hausse »}$ = Nombre de réponses « en hausse »,

$N \text{ « stable »}$ = Nombre de réponses « stable »,

$N \text{ « en baisse »}$ = Nombre de réponses « en baisse »,

et $N = N \text{ « en hausse »} + N \text{ « stable »} + N \text{ « en baisse »}$.

¹⁷ Dans le cas des délais de paiement, le solde d'opinion correspond classiquement à la différence entre, d'une part, la proportion de réponses envisageant une hausse et, d'autre part, celle envisageant une baisse. Une hausse du solde relatif aux délais de paiement traduit donc un allongement de ces délais.

¹⁸ Dans tout cet alinéa, afin d'alléger les notations, « en hausse » représente la réponse positive (en hausse, meilleure, s'allonger...), « stable » la réponse intermédiaire (stable, équivalente...) et « en baisse », la réponse négative (en baisse, plus mauvaise, se raccourcir...).

On calcule ensuite la moyenne pondérée appelée *solde d'opinion* (cf. troisième partie, section 3.1.1). Celle-ci correspond à l'opinion résumée des entreprises de la strate, calculée comme la différence entre la proportion des entreprises ayant une opinion positive de la situation et la proportion de celles qui en ont une opinion négative :

$$SOLDE = \% HAUSSE - \% BAISSE \quad (7)$$

Pour les perspectives générales d'activité, on calcule en outre un solde d'opinion non pondéré :

$$SOLDE_{\text{non pondéré}} = \% HAUSSE_{\text{non pondéré}} - \% BAISSE_{\text{non pondéré}} \quad (8)$$

On calcule enfin le pourcentage pondéré de non-réponses pour chaque question et pour chaque strate (relativement à l'ensemble des entreprises répondantes à l'enquête dans la strate) :

$$\% NR = 100 \times \left(1 - \frac{NP}{NPT} \right) \quad (9)$$

avec :

NP = Somme des pondérations des entreprises répondantes de la strate à la question considérée (définie *supra*),

NPT = Somme des pondérations des entreprises répondantes à l'enquête dans la strate, soit :

$$NPT = NP + NPN$$

où NPN = Somme des pondérations des non-réponses à la question dans la strate (parmi les entreprises répondantes à l'enquête dans la strate).

2.4.2.1.2 - Questions bimodales

Il s'agit des questions qualitatives à deux modalités (oui/non, cochée/non cochée).

Sont ici calculés les pourcentages pondérés de réponses « oui » et « non » pour chaque strate élémentaire :

$$\% OUI = 100 \times \frac{NP \ll \text{oui} \gg}{NP} \quad (10)$$

$$\% NON = 100 \times \frac{NP \ll \text{non} \gg}{NP} \quad (11)$$

avec :

$NP \ll \text{oui} \gg$ = Somme des pondérations des réponses « oui »,

$NP \ll \text{non} \gg$ = Somme des pondérations des réponses « non »,

et $NP = NP \ll \text{oui} \gg + NP \ll \text{non} \gg$.

2.4.2.2. - Agrégation secondaire des résultats

L'agrégation secondaire consiste, pour une question et un niveau d'agrégation donnés, à calculer une moyenne pondérée des soldes d'opinion élémentaires (issus de l'agrégation primaire), afin de tenir compte de l'importance relative de la strate dans l'ensemble du champ de l'enquête. Le solde d'opinion agrégé peut ainsi s'écrire :

$$SOLDE_{\substack{\text{agrégé.} \\ \text{niveau d'agrégation } J}} = \frac{\sum_{i \in \text{niveau d'agrégation } J} \beta_i \times SOLDE_i}{\sum_{i \in J} \beta_i}$$

où i est la strate élémentaire, β_i est le coefficient de redressement et $SOLDE_i$ le solde d'opinion issu de l'agrégation primaire.

Les coefficients de redressement β_i proviennent des données issues de la dernière Enquête Annuelle d'Entreprise (EAE) connue, prochainement de l'ESA. Les questions sont toutes pondérées par le chiffre d'affaires.

La même procédure est appliquée aux pourcentages de réponses à chacune des modalités « en hausse », « stable » et « en baisse ».

2.5. - La correction des variations saisonnières (CVS)

Les évolutions des réponses d'un mois sur l'autre peuvent s'avérer être, pour certaines entreprises, davantage liées à des variations saisonnières que conjoncturelles. Or, les soldes d'opinion, indicateurs résumés de ces réponses, visent à transcrire l'évolution conjoncturelle du secteur. Aussi est-il nécessaire de retirer l'information saisonnière contenue dans les séries obtenues à partir de l'agrégation des réponses individuelles, pour ne retenir que l'information d'ordre conjoncturel. C'est l'objet de la correction des variations saisonnières.

2.5.1. - Méthode générale

Des corrections de variations saisonnières d'un certain nombre de soldes d'opinion sont effectuées une fois par an¹⁹. Elles sont calculées annuellement par le programme (de type X11-ARIMA, X12-ARIMA). Pour obtenir une série désaisonnalisée à partir d'une série brute, on soustrait la série de coefficients saisonniers de la série brute :

$$VAR_{CVS}^q = VAR_{BRUTE}^q - VAR_{CS}^q$$

où VAR représente la série traitée, q la question et CS le coefficient saisonnier.

Les séries de coefficients saisonniers sont mises à jour chaque année pour tenir compte de l'influence des derniers points sur la valeur des coefficients saisonniers. Par conséquent, les séries désaisonnalisées sont recalculées annuellement. Toutes les questions sont corrigées des variations saisonnières. Des coefficients saisonniers de valeur nulle correspondent à des séries ayant été soumises à la procédure de correction des variations saisonnières, mais ne présentant pas de saisonnalité.

¹⁹ Les pourcentages de réponses « en hausse », « stable », et « en baisse », ainsi que les pourcentages de non-réponses et de réponses interpolées ne sont pas désaisonnalisés.

2.5.2. - Mode de report des coefficients saisonniers sur les observations récentes

Le mode de report des coefficients saisonniers est le suivant : en attendant la prochaine mise à jour des coefficients CVS, on utilise les derniers coefficients CVS calculés. En d'autres termes, on récupère la dernière valeur immédiatement existante du coefficient pour le trimestre T :

$$VAR_{CVS}^q(T, A) = VAR_{BRUTE}^q(T, A) - VAR_{CS}^q(T, A - i)$$

où T est un trimestre de l'année courante A et $A-i$ l'année du dernier coefficient calculé pour le trimestre T . En pratique, i prend la valeur 1, voire la valeur 2 si deux campagnes de CVS successives sont séparées d'un peu plus d'un an.

2.6. - Le stockage des données agrégées

Sont stockées à la fois les séries brutes et CVS. Les séries relatives aux perspectives générales d'activité sont stockées sous leurs formes pondérée et non pondérée.

