

---

***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 résulte d'un sondage stratifié sans remise à partir de l'Enquête Annuelle d'Entreprise (EAE). 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'industrie 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. Lors de ce renouvellement, on détermine un échantillon théorique de 3 000 entreprises, selon un sondage stratifié suivant un croisement « taille de l'entreprise » en termes d'effectifs salariés et « secteur d'activité ».

Dans un second temps, on détermine le nombre d'entreprises à tirer pour compléter l'échantillon de l'année précédente une fois toilé (c'est-à-dire après suppression des entreprises disparues, restructurées, etc.). Les nouvelles entreprises sont choisies par tirage aléatoire simple sans remise dans chaque strate<sup>16</sup>.

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.

<sup>16</sup> Rappel : ce mode de construction de l'échantillon ne s'effectue que pour les entreprises de tranches 1 à 5 incluses. La tranche 6 (cinq cents salariés et plus) est tirée de façon exhaustive.

## **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 vers le 27 du mois  $m-1$ . Les réponses doivent parvenir à l'Insee 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.

### ***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, ainsi que sur les évolutions de ces chiffres d'affaires d'une année à 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.

## **2.3. - 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 à choix multiple et montant des investissements.

## **2.4. - Le dépouillement**

Comme indiqué précédemment, seules certaines questions conjoncturelles sont effectivement posées tous les mois. D'autres ne sont posées que tous les trois mois.

Le dépouillement d'une question conjoncturelle s'effectue en deux étapes. On procède d'abord à l'agrégation des réponses au niveau de chaque strate élémentaire (voir *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), 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  $m$  et une seconde fois au moment de la publication de l'enquête suivante  $m+1$ . Ce procédé est motivé par deux raisons essentielles :

- l'application de la méthode de l'échantillon constant à certaines questions (Cf. alinéa 2.4.2) ;
- 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 taille<sup>1</sup>. On distingue 4 secteurs différents et 6 tranches de taille : il existe donc  $4 \times 6 = 24$  strates élémentaires.

### **2.4.2. - L'échantillon constant**

L'interprétation des résultats d'une enquête est largement fondée sur leur comparaison avec ceux de l'enquête précédente. En effet, comme on l'expliquera plus en détail dans la partie 3, alinéa 3.1.1, les résultats de ce type d'enquête n'ont de sens qu'en évolution, et non en niveau. Compte tenu du caractère qualitatif de la plupart des questions et du nombre limité d'entreprises interrogées, il est préférable de mener une telle comparaison sur la base d'un même échantillon pour deux enquêtes successives (c'est-à-dire « en raisonnant à échantillon constant »). On s'assure ainsi qu'une évolution des résultats entre deux enquêtes reflète exclusivement un changement dans les réponses des entreprises et non une évolution de structure de l'échantillon.

Cependant, pour une question donnée, prendre exclusivement en compte les opinions des entreprises ayant répondu à deux enquêtes successives aurait l'inconvénient de négliger les informations potentiellement instructives données par les entreprises ayant répondu à l'une seulement de ces deux enquêtes. En tout état de cause, en raison de la taille limitée des échantillons, on préfère prendre en compte le plus de réponses possible. Ceci implique plusieurs choix méthodologiques. D'une part, on constitue un échantillon constant par question plutôt que par enquête : ceci assure une couverture maximale de réponses à chaque question. D'autre part, dans le cas de réponses partielles, on est amené à définir des procédures de complèment. Ces procédures, très simples, consistent à imputer la réponse manquante d'une entreprise à une question donnée sur la base des réponses les plus récentes de l'entreprise à cette question.

La méthode dite de « l'échantillon constant », qui décrit l'ensemble de ces traitements (constitution de l'échantillon constant et traitement des non-réponses), est détaillée dans les alinéas suivants.

#### **2.4.2.1. - Les principes généraux de la constitution de l'échantillon constant**

La méthode de l'échantillon constant propose un mode de traitement de la non-réponse relativement simple, fondé sur un ensemble de règles de report des réponses antérieures des entreprises non répondantes. Elle permet de n'attribuer une évolution des résultats qu'à une modification de l'opinion des entreprises. Pour une question donnée, les réponses prises en compte dans les résultats sont celles des entreprises qui appartiennent à l'échantillon constant.

Les principes généraux de la méthode de l'échantillon constant sont les suivants :

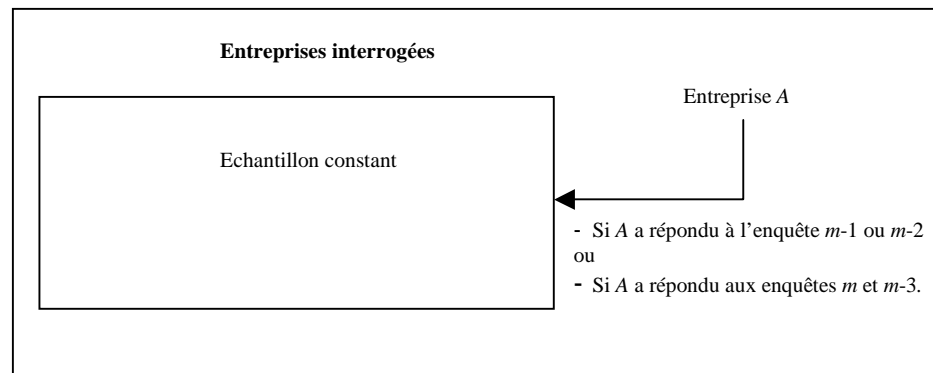
- l'échantillon constant est constitué question par question ;
- toutes les questions mensuelles, sauf celle relative aux prix, sont dépouillées selon la méthode de l'échantillon constant. Aucune question à périodicité trimestrielle n'est dépouillée selon cette méthode ;

<sup>1</sup> Rappel de la définition des tranches de taille, spécifique à l'enquête : - tranche 1 : 11 à 19 salariés - tranche 2 : 20 à 49 salariés - tranche 3 : 50 à 99 salariés - tranche 4 : 100 à 199 salariés - tranche 5 : 200 à 499 salariés - tranche 6 : 500 salariés et plus.

- une entreprise fait partie de l'échantillon constant si elle a répondu à l'enquête  $m-1$  ou  $m-2$  ou bien si elle a répondu aux enquêtes  $m$  (enquête courante) et  $m-3$  (Cf. alinéa 2.4.2.2, tableau 1). Le schéma suivant décrit cette règle.

Question  $Q$

Enquête  $m$



### 2.4.2.2. - Détail du mode de complèment des réponses partielles

Les deux tableaux ci-après explicitent la méthode de l'échantillon constant pour toutes les questions mensuelles sauf celles relatives au bloc « capacités de production »<sup>2</sup>.

<sup>2</sup> Les questions du bloc « capacités de production », n'aboutissant pas à des soldes d'opinion, font l'objet d'un traitement plus complexe qu'il n'est pas nécessaire d'explicitier ici.

**Tableau 1 : Décision d'inclusion dans l'échantillon constant et mode de complètemnt des réponses partielles pour une question  $Q$  donnée**

L'entreprise $A$ a répondu ( $R$ ) ou non ( $NR$ ) à la question $Q$ lors des 4 dernières enquêtes				Constitution de l'échantillon constant pour la question $Q$ lors de l'enquête $m$ : Traitements effectués pour l'entreprise $A$	
Enquête $m-3$	Enquête $m-2$	Enquête $m-1$	Enquête $m$	Inclusion ou non de l'entreprise $A$ dans l'échantillon constant	Prise en compte des réponses de l'entreprise $A$ et mode de complètemnt de ses non-réponses
$R$ ou $NR^3$	$R$ ou $NR$	$R$	$R$	<b>OUI</b>	Enquêtes $m-1$ et $m$ : les réponses de l'entreprise $A$ sont prises en compte dans les résultats.
$R$ ou $NR$	$R$ ou $NR$	$R$	$NR$	<b>OUI</b>	Enquête $m-1$ : la réponse de l'entreprise $A$ est prise en compte dans les résultats. Complètemnt de la question à l'enquête $m$ : la réponse de l'entreprise $A$ à l'enquête $m-1$ est reconduite à l'identique pour l'enquête $m$ .
$R$ ou $NR$	$R$	$NR$	$R$	<b>OUI</b>	Enquête $m-1$ : le traitement spécifique qui est appliqué est explicité dans le tableau 2 <i>infra</i> . Enquête $m$ : la réponse de l'entreprise $A$ est prise en compte dans les résultats.
$R$ ou $NR$	$R$	$NR$	$NR$	<b>OUI</b>	Complètemnt de la question aux enquêtes $m$ et $m-1$ : la réponse de l'entreprise $A$ à l'enquête $m-2$ est reconduite à l'identique pour les enquêtes $m-1$ et $m$ .
$R$	$NR$	$NR$	$R$	<b>OUI</b>	Enquête $m-1$ : le traitement spécifique qui est appliqué est explicité dans le tableau 2 <i>infra</i> . Enquête $m$ : la réponse de l'entreprise $A$ est prise en compte dans les résultats.
$NR$	$NR$	$NR$	$R$	NON	
$R$ ou $NR$	$NR$	$NR$	$NR$	NON	

Il arrive ainsi qu'une entreprise soit exclue de l'échantillon constant pour une question à laquelle elle a répondu lors de l'enquête courante  $m$  (ceci correspond à l'avant-dernière configuration explicitée dans le tableau 1). Dans ce cas, sa réponse à cette question n'est pas prise en compte dans le dépouillement des premiers résultats (c'est-à-dire lors de la publication de l'enquête en cours  $m$ ). Cependant, en vertu des règles décrites dans le tableau 1, cette réponse est réintégrée lors du dépouillement des résultats définitifs (c'est-à-dire lors de la publication de l'enquête suivante  $m+1$ ).

Dans les configurations ( $R$  ou  $NR$ ,  $R$ ,  $NR$ ,  $R$ ) ou ( $R$ ,  $NR$ ,  $NR$ ,  $R$ ), le mode de complètemnt appliqué pour l'enquête  $m-1$  diffère selon que les questions sont qualitatives (tableau 2) ou quantitatives.

<sup>3</sup> Nota : le symbole  $R$  indique que l'entreprise a répondu à la question considérée lors de l'enquête correspondante. Le symbole  $NR$  signifie qu'elle n'a pas donné de réponse à cette question au moment du dépouillement de l'enquête (questionnaire non parvenu ou réponse absente du questionnaire).

**Tableau 2 : Mode de complèment des réponses partielles de l'entreprise A  
dans les cas (R ou NR, R, NR, R) ou (R, NR, NR, R)**

A l'enquête <i>m-2</i> ou <i>m-3</i> <sup>4</sup> , l'entreprise A a répondu :	A l'enquête <i>m-1</i> , l'entreprise A n'a pas répondu :	A l'enquête <i>m</i> , l'entreprise A a répondu :	Lors de l'enquête <i>m-1</i> , la réponse suivante (« intermédiaire » entre les réponses aux enquêtes <i>m-2</i> ou <i>m-3</i> et <i>m</i> ) est reportée pour l'entreprise A :
(+)	NR	(+) ou (=)	(+)
(+)	NR	(-)	(=)
(=)	NR	(+), (=) ou (-)	(=)
(-)	NR	(+)	(=)
(-)	NR	(=) ou (-)	(-)

Lecture : (+) = réponse : en hausse, supérieur à la normal. (=) = réponse : stable, proche de la normale. (-) = réponse : en baisse, inférieur à la normale.

Pour les questions quantitatives, le mode de complèment des réponses de l'entreprise, dans les cas du tableau 2, est le suivant : calcul de la moyenne des deux réponses aux enquêtes *m-2* (ou *m-3* si l'entreprise n'a pas répondu à l'enquête *m-2*) et *m*.

Une entreprise qui n'a pas répondu à une question, mais dont la réponse est complétée selon l'une des modalités décrites dans les tableaux précédents, est dite « interpolée » pour cette question.

### 2.4.3. - Le dépouillement des questions

Dans toute cette section, sauf mention contraire, lorsque l'on parlera d'entreprises répondantes à une question, il s'agira des entreprises répondantes y compris les entreprises interpolées, c'est-à-dire après application de l'échantillon constant pour la question concernée. Par extension, lorsque l'on parlera d'entreprises répondantes, il s'agira de la réunion des entreprises répondantes à au moins une question, telles que définies ci-dessus.

#### 2.4.3.1. - 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, et sur le chiffre d'affaires (total, par type de travaux ou par type de clientèle) pour les autres questions.

##### 2.4.3.1.1 - Questions trimodales

Il s'agit de questions qualitatives, appelant une réponse positive (en hausse, supérieur à la normale, etc.), intermédiaire (stable, normal, etc.) ou négative (en baisse, inférieur à la normale, etc.).

<sup>4</sup> Dans le cas où l'entreprise a répondu aux enquêtes *m-2* et *m-3*, c'est la réponse la plus récente (*m-2*) qui est prise en compte.

Sont calculés les pourcentages pondérés de réponses « en hausse », « stable » et « en baisse » pour chaque question et chaque strate élémentaire<sup>5</sup> :

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

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

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

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

On calcule ensuite la moyenne pondérée appelée *solde d'opinion*<sup>6</sup>. 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 :

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

<sup>5</sup> Dans tout cet alinéa, afin d'alléger les notations, « en hausse » représente la réponse positive (en hausse, supérieur à la normale, ...), « stable » la réponse intermédiaire (stable, normal, ...) et « en baisse », la réponse négative (en baisse, inférieur à la normale, ...).

<sup>6</sup> Les soldes d'opinion sont interprétés avec plus de précision dans la troisième partie, section 3.1.1.



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

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

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).

Pour les perspectives générales d'activité, on calcule également un pourcentage non pondéré de non-réponses pour chaque strate (relativement à l'ensemble des entreprises répondantes à l'enquête dans la strate considérée) :

$$\% NR_{\text{non pondéré}} = 100 \times \left( 1 - \frac{N}{NT} \right) \quad (10)$$

avec :

$N$  = Nombre d'entreprises répondantes de la strate à la question considérée (défini *supra*),

$NT$  = Nombre d'entreprises répondantes à l'enquête dans la strate, soit :

$$NT = N + NN$$

où  $NN$  = Nombre d'entreprises non répondantes à la question considérée (parmi les entreprises répondantes à l'enquête dans la strate).

Enfin, pour toutes les questions pour lesquelles la procédure d'échantillon constant est appliquée, on calcule le pourcentage pondéré des réponses interpolées pour chaque strate (relativement à l'ensemble des entreprises répondantes à la question dans la strate après procédure d'échantillon constant) :

$$\% INT = 100 \times \frac{NPI}{NP} \quad (11)$$

avec :

$NPI$  = Somme des pondérations des entreprises interpolées de la strate pour la question considérée et  
 $NP$  = Somme des pondérations des entreprises répondantes à la question après application de l'échantillon constant dans la strate (définie *supra*), soit :

$$NP = NPR + NPI$$

où  $NPR$  = Somme des pondérations des vraies réponses à la question (interpolations exclues).

#### 2.4.3.1.2 - 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 :

$$\% OUI = 100 \times \frac{NP \text{ « oui »}}{NP} \quad (12)$$

$$\% NON = 100 \times \frac{NP \text{ « non »}}{NP} \quad (13)$$

avec :

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

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

et  $NP = NP \text{ « oui »} + NP \text{ « non »}$ .

#### **2.4.3.1.3 - Questions quantitatives**

Il s'agit des questions appelant une réponse numérique (un nombre de mois pour la question sur le travail assuré par le carnet de commandes, un pourcentage pour la question portant sur l'accroissement moyen possible de la production). On calcule alors une moyenne des réponses pondérée par le chiffre d'affaires total :

$$Q = \frac{SPR}{SP} \quad (14)$$

avec :

$SPR$  = Somme pondérée des réponses,

$SP$  = Somme des pondérations des répondants.

#### **2.4.3.2. - Agrégation secondaire des résultats**

Pour une question donnée, le solde d'opinion au niveau d'un regroupement de strates élémentaires est calculé comme la moyenne pondérée des soldes d'opinion élémentaires. La pondération permet de tenir compte de l'importance relative de la strate dans l'ensemble du champ de l'enquête.

Les coefficients de redressement proviennent de la dernière enquête annuelle d'entreprise connue. 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 », ainsi qu'au pourcentage de réponses interpolées.

## 2.5. - La correction des variations saisonnières (CVS)

### 2.5.1. - Méthode générale

Des corrections de variations saisonnières d'un certain nombre de soldes d'opinion sont effectuées une fois par an<sup>7</sup>. Elles sont calculées annuellement par le programme *X11-ARIMA*.

Pour produire 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.

Toutes les questions conjoncturelles font l'objet de la campagne de CVS.

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

Le mode de report des coefficients saisonniers est le suivant : en cas de non 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 mois *M* :

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

où *M* est un mois de l'année courante *A* et *A-i* l'année du dernier coefficient calculé pour le mois *M*. 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 stockés à la fois les séries brutes et CVS, les résultats issus de l'échantillon constant et les résultats issus de l'échantillon courant. Les séries relatives aux perspectives générales d'activité sont stockées sous leurs formes pondérée et non pondérée.

<sup>7</sup> 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.