PART TWO: SURVEY EXECUTION

PART TWO: SURVEY EXECUTION

Survey execution comprises several phases involving the team responsible for the survey at the national level (located at the INSEE Department of Short-Term Economic Analysis) and a team of computer specialists (based at several INSEE National Computing Centres).

The stages of survey execution in the narrow sense (excluding dissemination, interpretation, and application of results) are the following:

- sample formation and replacement;
- data collection and keying;
- preparation and editing of individual data;
- survey processing;
- seasonal adjustment;
- storage of aggregate data.

2.1. - Sample formation and replacement

The initial sample of enterprises to be surveyed is chosen by means of stratified sampling without replacement in the database of the latest Annual Enterprise Survey (Enquête Annuelle d'Entreprise: EAE) at the time of the survey launch. The EAE will shortly be replaced by the Annual Industry Survey (Enquête Sectorielle Annuelle: ESA). Most of the sample—which is an enterprise panel—is maintained from one Building Crafts Survey to the next. However, the sample is regularly updated owing to economic developments (death or restructuring of enterprises) and the need to preserve a sample of adequate size.

The Building Crafts Survey sample is thus partially replaced once a year, using identical procedures from year to year. The sampling frame is the latest available EAE (ESA from 2010 onwards). For the replacement, we calculate a theoretical sample of 1,500 enterprises, using a stratified sampling technique that combines two criteria: (1) "enterprise size" measured by workforce size (number of salaried employees) and (2) "activity sector". The number of enterprises in each stratum of the sample is proportional to the stratum's weight in terms of turnover in the total building crafts sector.

Secondly, we determine the number of enterprises to be picked in order to fill out the previous year's sample after "cleaning", i.e., removal of enterprises that have gone out of business, have been restructured, are poor respondents, and so on. The new enterprises are chosen by simple random sampling without replacement in each stratum.

Between two replacement operations, INSEE performs ongoing sample maintenance by procedures such as (1) continuously updating the addresses of surveyed enterprises and the correspondents in each enterprise, and (2) tracking the restructuring and sale of enterprises.

For more details on the sample, see the summary technical description of the Survey in Appendix 1.

2.2. - Data collection and keying

2.2.1. - Data collection

The questionnaires for the month *m* Survey are sent to sampled enterprises around the 27th of month *m*-1. The enterprises are asked to return their responses by around the 7th of the Survey month to INSEE, which captures the questionnaires.

In each Survey, follow-up questionnaires are sent to enterprises that have not responded by the return date printed on the initial questionnaire.

In January 2008, INSEE introduced a new collection method in parallel with the standard postal procedure: an electronic questionnaire available on http://conjoncture.entreprises.insee.fr. Enterprises can choose between returning the paper form or responding online.

2.2.2. - Keying and editing individual data

The data are captured by INSEE's Business Surveys Division as the questionnaires arrive.

The Division performs several data edits:

- consistency edits on structural data: these focus on consistency of total turnover and turnover by project category, workforce size, and year-to-year changes in these structural data;
- edits on the codings authorised for keying responses to the “short-term” questions: specific values are allowed depending on enterprise responses;
- classification edits: a keying classification table has been introduced in the keying application; it consists of the classification of economic activities.

2.2.3. - Online collection

Since January 2008, small businesses in the building industry can respond to the Building Crafts Survey online. They simply need to connect to our website and identify themselves with their access codes and passwords supplied by INSEE. They can then answer the questionnaire exactly as they would on paper.

The website offers explanations and services to facilitate response, as well as enhanced information feedback (§3.2.2). We are now using this new collection method in nearly all short-term business surveys.

When enterprises respond online, their responses are checked. Questionnaires that meet all the tests are entered directly into the individual-data files. Questionnaires displaying anomalies are dealt with manually.

2.3. - Organisation of individual data

The individual data are stored in a database common to all INSEE business surveys (Industry, Construction [i.e., Building/Public Works], Services, etc.), which contains:

- information describing the enterprises, such as SIREN numbers, names, and addresses;
- structural data for each enterprise: turnover, number of employees, the enterprise’s NAF code, and turnover by project category;
- short-term data: responses to qualitative questions.

2.4. - Processing

Processing short-term questions consists in computing aggregated levels from individual data. This involves two stages: (1) aggregation of responses for each primary stratum (see §2.4.1); (2) aggregation of results obtained in (1), so as to reflect the structure of the Survey field of coverage.

The first processing stage (primary aggregation) concerns only Survey data, i.e., responses to short-term questions and structural data (which serve as weighting coefficients). The second stage draws on external data, which serve as imputation coefficients. The external data are currently taken from the Annual Enterprise Survey (EAE); in the near future, they will be drawn from the Annual Industry Survey (ESA).

All Survey questions are processed twice: first for the publication of the current Survey, then for the publication of the following Survey. This is done to incorporate late responses.

2.4.1. - Stratification

A primary stratum is defined by the combination of (1) an activity sector (industry) grouping several categories at NAF 700 level and (2) a size bracket[16]. There are four different sectors (§1.4) and three size brackets: the number of primary strata is therefore $4 \times 3 = 12$.

2.4.2. - Question processing and determination of balances of opinion

2.4.2.1. - Primary aggregation of results

Primary aggregation consists in calculating the weighted averages of individual responses within each primary stratum. The weighting coefficient is a datum supplied by the enterprise. The coefficient is based on:

- workforce size for questions on past and expected workforce size and on hiring problems;
- turnover (total, by project category, or by customer category) for the other questions.

2.4.2.1.1. - Three-choice questions

These are qualitative questions, calling for a response that is either positive (increasing, better, etc.), intermediate (unchanged, normal, etc.) or negative (decreasing, worse, etc.)[17].

We then calculate the weighted percentages of "increasing", "unchanged", and "decreasing" responses for each question and each primary stratum[18]:

$$(1) \qquad \%\ \text{INCREASE} = 100 \times \frac{\text{“increasing” } SW}{SW}$$

$$(2) \qquad \%\ \text{UNCHANGED} = 100 \times \frac{\text{“unchanged” } SW}{SW}$$

$$(3) \qquad \%\ \text{DECREASE} = 100 \times \frac{\text{“decreasing” } SW}{SW}$$

[16] Reminder of brackets specific to the Building Crafts Survey: bracket 0 = 0 employees; bracket 1 = 1-5 employees; bracket 2 = 6-10 employees.

[17] For the question regarding settlement times, the balance of opinion, by convention, is defined as the difference between the proportions of responses indicating longer times and responses indicating shorter times. An increase in the balance of opinion on settlement times therefore indicates that customers are taking longer to pay.

[18] Throughout this section, to simplify notations, "increasing" denotes the positive response (increasing, better, getting longer [for settlement times], etc.), "unchanged" the "intermediate" response (unchanged, comparable, etc.), and "decreasing" the negative response (decreasing, worse, getting shorter [for settlement times], etc.).

With:

"increasing" SW = Sum of weights of "increasing" responses

"unchanged" SW = Sum of weights of "unchanged" responses

"decreasing" SW = Sum of weights of "decreasing" responses

SW = "increasing" SW + "unchanged" SW + "decreasing" SW.

For the general business outlook, we also determine the unweighted percentages of "increasing", "unchanged", and "decreasing" responses for each primary stratum:

$$(4) \qquad \%\ \text{INCREASE}_{\text{unweighted}} = 100 \times \frac{\text{"increasing" } N}{N}$$

$$(5) \qquad \%\ \text{UNCHANGED}_{\text{unweighted}} = 100 \times \frac{\text{"unchanged" } N}{N}$$

$$(6) \qquad \%\ \text{DECREASE}_{\text{unweighted}} = 100 \times \frac{\text{"decreasing" } N}{N}$$

With:

"increasing" N = Number of "increasing" responses

"unchanged" N = Number of "unchanged" responses

"decreasing" N = Number of "decreasing" responses

N = "increasing" N + "unchanged" N + "decreasing" N.

We then determine the weighted average called *balance of opinion* (§3.1.1). This is the summarised opinion of enterprises in the stratum, defined as the difference between the proportion of enterprises voicing a positive opinion of the situation and the proportion of those with a negative opinion:

$$(7) \qquad BALANCE = \%\ INCREASE - \%\ DECREASE$$

For the general business outlook, we also compute an unweighted balance of opinion:

$$(8) \qquad BALANCE_{unweighted} = \%\ INCREASE_{unweighted} - \%\ DECREASE_{unweighted}$$

Lastly, we calculate the weighted percentage of non-responses for each question and for each stratum (relative to total Survey respondents in the stratum):

$$(9) \qquad \%\ \text{NR} = 100 \times \left(1 - \frac{SW}{TSW}\right)$$

With:

SW = Sum of weights of stratum respondents to particular question (defined above)

TSW = Total sum of weights of Survey respondents in stratum:

$TSW = SW + SWN$

where SWN = Sum of weights of non-responses to question in stratum (among Survey respondents in stratum).

2.4.2.1.2 - Two-choice questions

These are qualitative questions with two choices: yes/no, box ticked/unticked.
We calculate the weighted percentages of "yes" and "no" responses for each primary stratum:

$$\% \text{ YES} = 100 \times \frac{SW \text{ "yes"}}{SW} \quad (10)$$

$$\% \text{ NO} = 100 \times \frac{SW \text{ "yes"}}{SW} \quad (11)$$

With:

SW *"yes"* = Sum of weights of "yes" responses

SW *"no"* = Sum of weights of "no" responses

$SW = SW$ *"yes"* $+ SW$ *"no"*.

2.4.2.2. - Secondary aggregation of results

Secondary aggregation consists in determining, for a given question and aggregation level, the weighted average of the primary balances of opinion, which are obtained from the primary aggregation. This procedure enables us to make allowance for the stratum's size relative to the total Survey field. We can thus write the aggregated balance of opinion as follows:

$$BALANCE_{\substack{\text{aggregated},\\ \text{aggregation level } J}} = \frac{\sum_{i \in \text{ aggregation level } J} \beta_i \times BALANCE_i}{\sum_{i \in J} \beta_i}$$

where i is the primary stratum, β_i the imputation coefficient and $BALANCE_i$ the balance of opinion obtained from the primary aggregation.

The imputation coefficients β_i are taken from the latest available Annual Enterprise Survey (EAE); in the near future, the source will be the Annual Industry Survey (ESA). All questions are weighted by turnover.

The same procedure is applied to the percentages of "increasing", "unchanged", and "decreasing" responses.

2.5. - Seasonal adjustment (SA)

For some enterprises, monthly changes in survey responses may be due to seasonal fluctuations more than to business-cycle factors. Balances of opinion, which are summarised indicators of responses, aim to track short-term economic conditions in the sector. We must therefore remove the seasonal information contained in the

series obtained by aggregating individual responses, leaving only the information relating to the business cycle. That is the purpose of seasonal adjustment.

2.5.1. - General method

Selected balances of opinion are seasonally adjusted once a year[19] using *X*11-*ARIMA* and *X*12-*ARIMA* software. To produce a seasonally adjusted series from a raw series, we subtract the seasonal-coefficient series from the raw series:

$$VAR_{SA}^{q} = VAR_{RAW}^{q} - VAR_{SC}^{q}$$

where *VAR* is the series processed, *q* the question, and *SC* the seasonal coefficient.

The seasonal-coefficient series are updated annually to reflect the influence of the latest data points on seasonal-coefficient values. We therefore recalculate seasonally adjusted series annually. All questions are seasonally adjusted. Zero-value seasonal coefficients apply to series that have been seasonally adjusted but display no seasonal pattern.

2.5.2. - Method for applying seasonal coefficients to recent observations

Seasonal (*SA*) coefficients are carried forward as follows: until the *SA* coefficients are updated, we use the latest computed *SA* coefficients. In other words, we recapture the latest immediately existing value of the coefficient for quarter *T*:

$$VAR_{SA}^{q}(Q,Y) = VAR_{RAW}^{q}(Q,Y) - VAR_{SC}^{q}(Q,Y-i)$$

where *Q* is a quarter of the current year *Y* and *Y-i* the year of the latest coefficient calculated for quarter *Q*. In practice, *i* takes the value 1, or even the value 2 if two successive *SA* procedures are separated by slightly over a year.

2.6. - Storage of aggregated data

Raw and seasonally adjusted series are stored. The series on the general business outlook are stored in their weighted and unweighted forms.

[19] The percentages of "increasing", "unchanged", and "decreasing" responses are not seasonally adjusted; nor are the percentages of non-responses and interpolated responses.