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 Annual Enterprise Survey (Enquête Annuelle d'Entreprise: EAE). Most of the sample—which is an enterprise panel—is maintained from one Building Survey to the next. However, owing to economic developments (death or restructuring of enterprises) and the need to preserve a sample of adequate size, the sample is regularly updated.

The Building Survey sample is thus partially replaced once a year, using identical procedures from year to year. The sampling frame is the latest available Annual Enterprise Survey. For the replacement, we calculate a theoretical sample of 3,000 enterprises, using a stratified sampling technique that combines two criteria: (1) "enterprise size" measured by workforce size (number of salaried employees) and (2) "sector".

Secondly, we determine the number of enterprises to be picked to fill out the previous year's sample after "cleaning", i.e., removal of enterprises that have gone out of business, been restructured, and so on. The new enterprises are chosen by simple random sampling without replacement in each stratum[1].

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

More details on samples are provided in the summary technical description of the Survey in Appendix 1.

---

[1] Reminder: this sample-building method is used only for enterprises in size brackets 1-5. In bracket 6 (500 or more employees), all units are sampled.

2.2. - Data collection and keying

2.2.1. - Data collection

The questionnaires are sent to the enterprises in the sample around the 27th of month m-1. The responses must reach the Institute by around the 7th of the Survey month.

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

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, as well as on their year-to-year changes;

- edits on the codings allowed 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.

2.3. - 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, such as turnover, number of employees, the enterprise's NAF code;

- short-term data: responses to multiple-choice questions and investment figures.

2.4. - Processing

As noted in §1.5, only some short-term questions are actually asked every month. Others are asked only every three months.

The short-term questions are processed in 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., the responses to the short-term questions and the structural data, which serve as weighting coefficients. The second stage draws on external data (from the Annual Enterprise Survey), which serve as imputation coefficients.

All Survey questions are processed twice: once for the publication of the Survey for the current month m, and again for the publication of the following m+1 Survey. There are two basic reasons for this procedure:
- the application of the constant-sample method, described in §2.4.2;
- the incorporation of late responses.

2.4.1. - Stratification

A primary stratum is defined by an activity sector (industry) grouping few categories at the NAF 700 level combined with a size bracket[2]. There are 4 different sectors and 6 size brackets: the number of primary strata is therefore $4 \times 6 = 24$.

2.4.2. - The constant sample

The interpretation of the results of a Survey is largely based on their comparison with those of the previous Survey. This is because—as explained in greater detail in §3.1.1—the only meaningful aspect of the results of such Surveys is their variations, not their absolute levels. Given the qualitative nature of most of the questions and the small number of enterprises surveyed, it is preferable to perform such a comparison using an identical sample for two successive Surveys, i.e., on a "constant-sample" basis. This ensures that a change of results between two Surveys will exclusively reflect a change in the enterprises' responses and not a change in the sample structure.

However, for a given question, if we were to take into account only the opinions of enterprises having responded to two successive Surveys, we would overlook the potentially instructive information provided by the enterprises having responded to only one of the two Surveys. In any event, because of the smallness of the samples, we prefer to include as many responses as possible. This implies several methodological choices. First, we construct a constant sample for each question rather than for an entire Survey: this offers maximum coverage of the responses to each question. Second, to handle partial responses, we define completion procedures. These are very simple, and consist in imputing an enterprise's missing response to a particular question from its most recent responses to the question.

The "constant sample" method encompasses all these operations, i.e., determination of the constant sample and processing of non-responses. It is described in the following sub-sections.

2.4.2.1. - General principles for determining constant sample

The constant-sample method is a fairly simple way to process non-response, based on a set of rules for "carrying forward" the earlier responses of non-respondent enterprises. Changes in results can thus be explained solely as the consequence of a shift in enterprise opinion. For a given question, the responses taken into account in the results are those of the enterprises in the constant sample.
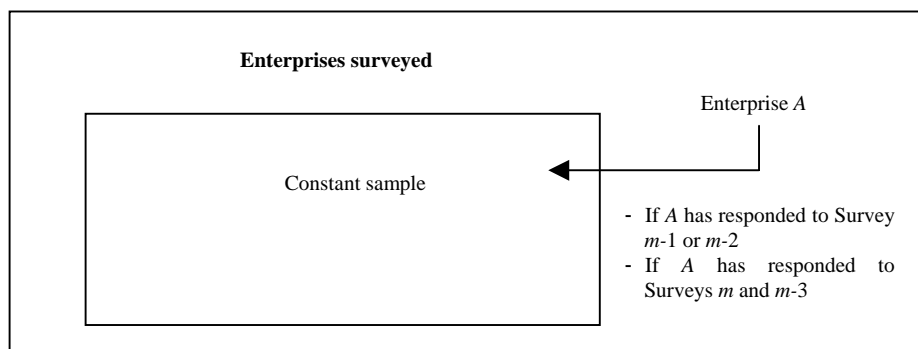
The general principles of the constant-sample method are as follows:

- The constant sample is formed for individual questions.

- The constant-sample method is applied to all monthly questions except the one on prices. No quarterly question is processed with the method.

- An enterprise is included in the constant sample if it has responded either (1) to the m-1 or m-2 Survey, or (2) to the m (current month) and m-3 Surveys (see §2.4.2.2., Table 1). The procedure is summarised in the following diagram.

---

[2] Reminder: bracket 1: 11-19 employees; bracket 2: 20-49 employees; bracket 3: 50-99 employees; bracket 4: 100-199 employees; bracket 5: 200-499 employees; bracket 6: 500 or more employees.

Enterprises surveyed

Enterprise *A*

Constant sample

- If *A* has responded to Survey *m*-1 or *m*-2
- If *A* has responded to Surveys *m* and *m*-3

2.4.2.2. - Details on method for completion of partial responses

The two tables below describe the specific procedure applied to all monthly questions except the "productive capacity" set[3].

Table 1: Decision to include an enterprise in the constant sample and
procedure for completing partial responses to a given question *Q*

Enterprise A has responded (R) or not (NR) to question Q in past 4 Surveys				Formation of constant sample for question *Q* in Survey *m*: Procedures for enterprise *A*	
Survey *m*-3	Survey *m*-2	Survey *m*-1	Survey *m*	Enterprise *A* included constant sample?	Processing of enterprise *A*'s responses and procedure for completing its non-responses
R or *NR*[4]	*R* or *NR*	*R*	*R*	**YES**	Surveys *m*-1 and *m*: enterprise *A*'s responses are included in results.
R or *NR*	*R* or *NR*	*R*	*NR*	**YES**	Survey *m*-1: enterprise *A*'s response is included in results. Completion of response to question in Survey *m*: enterprise *A*'s response to Survey *m*-1 is copied to Survey *m*.
R or *NR*	*R*	*NR*	*R*	**YES**	Survey *m*-1: specific procedure applied is described in Table 2 below. Survey *m*: enterprise *A*'s response is included in results.
R or *NR*	*R*	*NR*	*NR*	**YES**	Completion of response to question in Surveys *m* and *m*-1: enterprise *A*'s response to Survey *m*-2 is copied to Surveys *m*-1 and *m*.
R	*NR*	*NR*	*R*	**YES**	Survey *m*-1: specific procedure applied is described in Tables 2 and 3 below. Survey *m*: enterprise *A*'s response is included in results.
NR	*NR*	*NR*	*R*	NO	
R or *NR*	*NR*	*NR*	*NR*	NO	

---

[3] As the "productive capacity" questions do not yield balances of opinion, they are processed using a more complex procedure that need not be described here.

[4] The symbol *R* denotes that the enterprise has responded to this particular question in the Survey indicated. The symbol *NR* indicates that it did not respond to the question at the time the Survey was processed (questionnaire not received, or response missing from questionnaire).

An enterprise may therefore be excluded from the constant sample for a question to which it has responded in the current Survey *m* (this is the next-to-last configuration in Table 1). In this case, its response to the question is not included in the processing of preliminary results, i.e., for the publication of the current Survey *m*. However, under the rules set out in Table 1, this response is added back in the processing of final results, i.e., for the publication of the following Survey, *m*+1.

In the configurations (*R or NR, R, NR, R*) or (*R, NR, NR, R*), the completion method applied for Survey *m*-1 differs according to whether the questions are qualitative (Table 2) or quantitative.

Table 2: Procedure for completion of partial responses by enterprise *A*
in (*R or NR, R, NR, R*) or (*R, NR, NR, R*) configurations

In Survey *m*-2 or *m*-3[5], enterprise *A* responded:	In Survey *m*-1, enterprise *A* did not respond:	In Survey *m*, enterprise *A* responded:	In Survey *m*-1, the following response ("interpolated" between responses to Surveys *m*-2 or *m*-3 and *m*) is recorded for enterprise *A*:
(+)	*NR*	(+) or (=)	(+)
(+)	*NR*	(−)	(=)
(=)	*NR*	(+), (=) or (−)	(=)
(−)	*NR*	(+)	(=)
(−)	*NR*	(=) or (−)	(−)

Key:
(+) = "increasing", "above normal" response
(=) = "unchanged", "normal" response
(−) = "decreasing", "below normal" response.

For the quantitative questions, the completion procedure for enterprise responses in the case of Table 2 is as follows: we determine the mean of the two responses to Surveys *m*-2 (or *m*-3 if the enterprise has not responded to Survey *m*-2) and *m*.

An enterprise that has not responded to a question, but whose response is completed using one of the procedures described in the preceding tables, is said to be "interpolated" for this question.

2.4.3. - *Question processing*

In this entire section, "respondent enterprises" include interpolated enterprises, i.e., after application of the constant sample for the question concerned. By extension, "respondent enterprises" will be used to designate the aggregation of enterprises responding to at least one question, as defined above.

2.4.3.1. - 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 (1) workforce size for questions on past and expected workforce size and on hiring problems, and (2) turnover (total, by project category, or by customer category) for the other questions.

---

[5] If the enterprise has responded to Surveys *m*-2 and *m*-3, the more recent response (*m*-2) is the one taken into account.

2.4.3.1.1 - Three-choice questions

These are qualitative questions, calling for a response that is either positive (increasing, above normal, etc.), intermediate (unchanged, normal, etc.) or negative (decreasing, below normal, etc.).

We then calculate the weighted percentages of "increasing", "unchanged", and "decreasing" responses for each question and each primary stratum[6]:

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

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

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

With:

$\quad$ *"increasing" SW* = Sum of weights of "increasing" responses

$\quad$ *"unchanged" SW* = Sum of weights of "unchanged" responses

$\quad$ *"decreasing" SW* = Sum of weights of "decreasing" responses

$\quad$ *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:

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

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

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

With:

$\quad$ *"increasing" N* = Number of "increasing" responses

$\quad$ *"unchanged" N* = Number of "unchanged" responses

$\quad$ *"decreasing" N* = Number of "decreasing" responses

$\quad$ *N = "increasing" N + "unchanged" N + "decreasing" N.*

---

[6] Throughout §2.4.3, to simplify notations, "increasing" denotes the positive response (increasing, above normal, etc.), "unchanged" the "intermediate" response (unchanged, normal, etc.), and "decreasing" the negative response (decreasing, below normal, etc.).

We then determine the weighted average called *balance of opinion*[7]. 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)   *BALANCE = % INCREASE - % DECREASE*

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

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

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

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

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).

For the general business outlook, we also calculate an unweighted percentage of non-responses for each stratum (relative to total Survey respondents in stratum):

(10) $\% NR_{unweighted} = 100 \times \left(1 - \dfrac{N}{TN} \right)$

With:

N = Number of stratum respondents to particular question (defined above)
TN = Total number of Survey respondents in stratum:

$TN = N + NN$

where NN = Number of non-respondents to particular question (among Survey respondents in stratum).

Lastly, for all the questions to which the constant-sample method is applied, we calculate the weighted percentage of interpolated responses for each stratum (relative to total stratum respondents to question after constant-sample procedure):

(11) $\% INT = 100 \times \dfrac{SWI}{SW}$

With:

SWI = Sum of weights of interpolated enterprises in stratum for question examined
SW = Sum of weights of respondents to question after application of constant sample in stratum (defined above):

$SW = SWR + SWI$

where SWR = Sum of weights of actual responses to question (excluding interpolations).

---

[7] Balances of opinion are interpreted with greater precision in §3.1.1.

2.4.3.1.2. - 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:

$$(12) \qquad \% \; YES = 100 \times \frac{SW \text{ "yes"}}{SW}$$

$$(13) \qquad \% \; NO = 100 \times \frac{SW \text{ "yes"}}{SW}$$

With:

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

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

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

2.4.3.1.3 - Quantitative questions

These questions require a quantitative response: number of months for the question on work guaranteed by order books; a percentage for the question on the potential mean increase in production. We compute a mean response weighted by total turnover:

$$(14) \qquad Q = \frac{WSR}{SW}$$

With:

WSR = Weighted sum of responses

SW = Sum of respondents' weights.

2.4.3.2. - Secondary aggregation of results

For a given question, the balance of opinion for a primary-stratum grouping is the weighted average of the primary balances of opinion. Weighting enables us to make allowance for the stratum's size relative to the total Survey field.

The imputation coefficients are taken from the latest available Annual Enterprise Survey. All questions are weighted by turnover.

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

2.5. - Seasonal adjustment (SA)

2.5.1. - General method

Selected balances of opinion are seasonally adjusted once a year[8] using *X11-ARIMA* software.

To produce a seasonally adjusted series from a raw series, we subtract the seasonal-coefficient series from the raw series:

$$VAR^q_{SA} = VAR^q_{RAW} - VAR^q_{SC}$$

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

All qualitative questions are seasonally adjusted.

2.5.2. - Method for applying seasonal coefficients to recent observations

Seasonal coefficients are carried forward as follows: if the coefficients are not updated, we use the latest computed coefficients. In other words, we recapture the latest immediately existing value of the coefficient for month *M*:

$$VAR^q_{SA}(M, Y) = VAR^q_{RAW}(M, Y) - VAR^q_{SC}(M, Y - i)$$

where *M* is a month of the current year *Y* and *Y-i* the year of the latest coefficient calculated for month *M*. 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 series, seasonally adjusted series, results from the constant sample, and results from the current sample are stored. The series on the general business outlook are stored in their weighted and unweighted forms.

---

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