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;

- data storage.

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 available Annual Enterprise Survey (Enquête Annuelle d'Entreprise: EAE). Most of the sample—which is an enterprise panel—is maintained from one Industry 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 Survey sample is thus partially replaced once a year, using identical procedures from year to year. For the replacement, we calculate a theoretical sample of 4,000 enterprises, using a stratified sampling technique that combines two criteria: (1) "enterprise size" measured by workforce size (number of salaried employees) and (2) "sector"[1]. The number of enterprises to be included in the sample for each stratum is proportional to the stratum's share of total industry turnover (see Appendix 6). 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.

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

---

[1] Reminder: this sample-building method is used only for enterprises with fewer than 500 employees or with turnover of under €150 million. The other enterprises—with more than 500 employees or turnover of over €150 million—are sampled exhaustively.

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—which captures the questionnaire data—by around the 10th of the Survey month.

In each Survey, INSEE sends a follow-up questionnaires or email reminder to enterprises that have not responded by the return date printed on the initial questionnaire.

In November 2005, a new procedure was introduced as an alternative to the standard collection procedure by postal mail. It consists of an electronic questionnaire available online at http://conjoncture.entreprises.insee.fr. For each Survey, respondents can choose to answer on paper forms or on the Internet.

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, sales by product, and workforce size, as well as on year-to-year changes in these structural data;

- edits on responses to "short-term" questions.

Questionnaires returned online are edited and, when appropriate, incorporated directly into individual-data files. Questionnaires exhibiting anomalies are processed manually.

2.2.3. – Online data collection

In November 2005, INSEE offered business owners the option of responding to the Industry Survey online. Respondents need only connect to the Institute's website and identify themselves with their INSEE-supplied access code and password. They can then fill out the questionnaire exactly as they would on paper.

The website offers explanations and services to facilitate responses, as well as enhanced information feedback (see §3.2.2). INSEE will gradually extend this new collection method to all business surveys.

2.3. - Organisation of individual data

The individual data are stored in a database common to all INSEE business surveys, which contains:

- information describing the enterprises, such as SIREN (i.e., business register) numbers, names, and addresses;

- structural data, such as turnover, number of employees, and enterprise's principal activity;

- short-term data: responses to qualitative questions.

2.4. - Processing

Processing consists in determining aggregate levels from individual data. As noted in §1.5, only some short-term questions are actually asked every month. Others are asked only every quarter.

Aggregate values are computed in two stages: (1) aggregation of responses for each primary stratum (see §2.4.3); (2) aggregation of results obtained in (1), taking account the weight of each stratum in the Survey's scope of coverage.

The "enterprise" questions are aggregated using a "sector" approach. "Product" questions are aggregated using a "product" approach. Questions on raw materials are aggregated by applying a combination of "material used" and "user sector" criteria.

The first processing stage (primary aggregation) solely concerns Survey data: the structural data serve as weighting coefficients. The second processing stage draws on data from the Annual Enterprise Survey and the national accounts, 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 Survey in $m+1$. 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

- In the "sector" approach, a primary stratum is defined by a combination of an activity sector at the NES 114 level and a size bracket. There are 59 different sectors and 3 size brackets: the number of primary strata in the "sector" approach is therefore $59 \times 3 = 177$.

- In the "product" approach, a primary stratum is defined as an industry classified in NES 114. Under this "product" approach, the Survey covers 59 different primary strata.

- In the "material" approach, a primary stratum is defined as a combination of a user sector and a type of raw material used. User sectors and materials used are defined at the NES 36 level. There are 17 user sectors and 18 classes of raw materials, hence a total of $17 \times 18 = 306$ primary strata in the "material" approach.

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 meaningful aspect of the results of such Surveys is their variations, more than their 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.

- Most questions are processed using the constant-sample method[2].

- For each question processed with the constant-sample method, an enterprise is included in the constant sample only if it has responded to that particular question in at least two of the past three Surveys or if it has responded to the intermediate Survey. For the monthly questions, the Surveys comprise those of the current month m and two previous months m-1 and m-2. For the quarterly questions, we take the current month m and months m-3 and m-6. In the following discussion, the dates of the three surveys involved are written S, S-1, and S-2, for monthly and quarterly questions alike. The procedures are summarised in the following diagram.

---
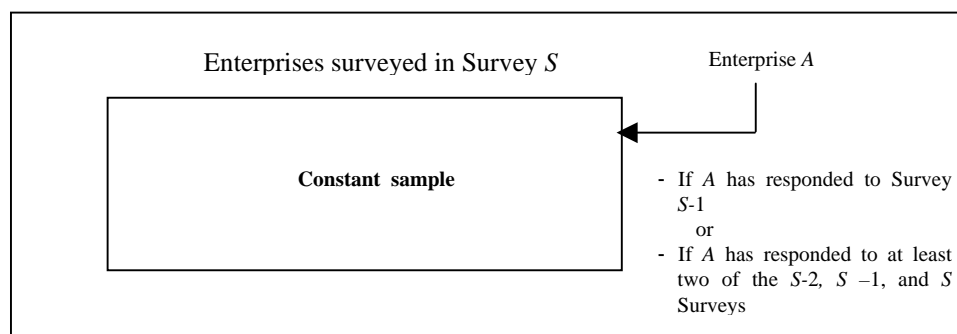
[2] The constant-sample method is not applied to questions on the general business outlook and to quarterly questions on prices and wage rates. This is because the carry-forward of earlier responses to the current Survey seems rather inappropriate for questions on prices (given their volatility) or questions asked at intervals exceeding three months (responses to the most recent survey would then refer to a period regarded as too remote). General sentiment is published in unweighted form, which substantially reduces the need for the constant-sample approach.

2.4.2.2. - Details on method for completion of partial responses

The two tables below describe the specific procedure applied under the constant-sample method for the Industry Survey.

Table 2: 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 3 Surveys			Formation of constant sample for question *Q* in Survey *S*: Procedures for enterprise *A*	
Survey *S* -2	Survey *S*-1	Survey *S*	Enterprise *A* included constant sample?	Processing of enterprise *A*'s responses and procedure for completing its non-responses
R[3]	*R*	*R*	**YES**	Surveys *S*-1 and *S*: enterprise *A*'s responses are included in results.
R	*R*	*NR*	**YES**	Survey *S*-1: enterprise *A*'s response is included in results. Completion of response to question in Survey *S*: enterprise *A*'s response to Survey *S*-1 is copied to Survey *S*.
R	*NR*	*R*	**YES**	Survey *S*-1: specific procedure applied is described in Table 3 below. Survey *S*: enterprise *A*'s response is included in results.
R	*NR*	*NR*	NO	Surveys *S*-1 and *S*: enterprise *A* is excluded from constant sample.
NR	*R*	*R*	**YES**	Surveys *S*-1 and *S*: : enterprise *A*'s responses to both Surveys are included in results.
NR	*R*	*NR*	**YES**	Survey *S*-1: enterprise *A*'s response is included in results. Completion of response to question in Survey *S*: enterprise *A*'s response to Survey *S*-1 is copied to Survey *S*.
NR	*NR*	*R*	NO	Surveys *S*-1 and *S*: enterprise *A* is excluded from constant sample.
NR	*NR*	*NR*	NO	Surveys *S*-1 and *S*: enterprise *A* is excluded from constant sample.

An enterprise may therefore be excluded from the constant sample for a question to which it has responded in the current Survey *S* (this is the next-to-last configuration in Table 2). 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 *S*. However, under the rules set out in Table 2, this response is added back in the processing of final results, i.e., for the publication of the following Survey, *S*+1.

---

[3] The symbol *R* denotes that the enterprise 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).

Table 3: Procedure for completion of partial responses by enterprise *A* in (*R*, *NR*, *R*) configuration

In Survey *S*-2, enterprise *A* responded:	In Survey *S*-1, enterprise *A* did not respond:	In Survey *S*, enterprise *A* responded:	In Survey *S*-1, the following response ("interpolated" between responses to Surveys *S*-2 and *S*) 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.

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.

All Survey questions are processed twice: once for the publication of the current Survey *S*, and again for the publication of the following Survey *S*+1. There are two basic reasons for this procedure:

- as noted above, it is required for the application of the constant-sample method to selected questions;

- it also allows the incorporation of late responses.

2.4.3. - *Question processing*

Question processing and the determination of balances of opinion differ according to the type of question. Questions on the enterprise as a whole are processed using a "sector" approach, questions on the enterprise's products are processed using a "product" approach, and questions on raw materials are processed using a "materials" approach.

2.4.3.1. - Processing "enterprise" questions

In this entire section, "enterprises responding a question" include, in addition to those which have actually answered this question, interpolated enterprises after application of the constant sample.

2.4.3.1.1. - *Primary aggregation of results*

Primary aggregation consists in calculating *balances of opinion* within each primary stratum. Primary strata are defined by combining an NES 114 activity sector and a workforce-size bracket. For each question, the *balance of opinion* is defined as the difference between the proportion of "individuals" ("individual" here denotes a "enterprise responding that question") voicing a positive opinion of the situation and the proportion of those with a negative opinion[4].

---

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

Let i a primary stratum*:*

$BALANCE_i = 100 \times (\% \ INCREASE_i - \% \ DECREASE_i)$

With:

$$\% \ INCREASE_i = \frac{\sum\limits_{\substack{k \in i \\ k \ responding \ "increasing"}} \alpha_k}{\sum\limits_{k \in i} \alpha_k}$$

$$\% \ DECREASE_i = \frac{\sum\limits_{\substack{k \in i \\ k \ responding \ "decreasing"}} \alpha_k}{\sum\limits_{k \in i} \alpha_k}$$

Where: $\sum\limits_{k \in i} \alpha_k = \sum\limits_{\substack{k \in i \\ k \ responding \ "increasing"}} \alpha_k + \sum\limits_{\substack{k \in i \\ k \ responding \ "unchanged"}} \alpha_k + \sum\limits_{\substack{k \in i \\ k \ responding \ "decreasing"}} \alpha_k$

And α_k represents the weight relating to the k enterprise.

The weight is an information supplied by the enterprise and depends on the question: for instance, the coefficient is based on workforce size for questions on past and expected workforce size and working time, for questions on hiring difficulties and on annual turnover for questions on productive capacity, cash-flow position, and factors limiting production. At last, the weighting coefficient is constant, equal to 1 for the general business outlook.

2.4.3.1.2. - Secondary aggregation of results

Secondary aggregation consists, for a given question and an aggregation level, in calculating a weighted average of the primary balances of opinion (computed from the primary aggregation), in order to make allowance for the primary stratum's size relative to this aggregation level. Thus, the aggregated balance of opinion can be defined as:

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

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

The imputation coefficients are taken from the latest available Annual Enterprise Survey and the national accounts. They change again according to the questions: workforce size (broken down by size bracket) is used for questions on past and expected changes in workforce size, working time, and hiring difficulties; value added by sector (broken down by size bracket) is used for questions on productive capacity, cash-flow position, and factors limiting production.

2.4.3.2. - Processing of "product" questions

Logic is the same one as in the previous part.

2.4.3.2.1. - Primary aggregation of results

Here, "individual" denotes the product of an enterprise. Each product is classified in the French Classification of Economic Activities (Nomenclature d'Activités Française: NAF) at the NAF 700 level. The primary stratum is a detailed activity sector at NES 114 level, which combines all products classified in the sector. At least, the weighting coefficient is the enterprise's annual export turnover of the product for questions on foreign demand, and the enterprise's total annual turnover of the product in France and abroad for the other questions on the enterprise's products.

2.4.3.2.2. - Secondary aggregation of results

Compared with the secondary aggregation of "enterprise" questions, only the imputation coefficient varies. Taken from the annual national-accounts data, it depends again on the question: the value of manufactured-goods exports is used to impute foreign-demand variables, while the value added by sector is used to impute the other activity variables concerning the enterprise's products.

2.4.3.3. - Processing of "materials" questions

Here again, the nature of the processing is the same.

2.4.3.3.1. - Primary aggregation of results

"Individual" here denotes a raw material used by the enterprise in its production process. Each raw material is classified in the Nomenclature Economique de Synthèse at NES 36 level and according to the sector (at NES 36 level) of the user enterprise. The primary stratum is thus defined by a combination of a user sector and a type of raw material used. The weighting coefficient is the enterprise's annual purchases of the material concerned.

2.4.3.3.2. - Secondary aggregation of results

In this case, the weights used to impute the primary balances of opinion are taken from the annual national-accounts data: they consist of the cross-tabulation (materials used × user industry) of intermediate consumption at NES 36 level.

2.5. - Seasonal adjustment (SA)

For some enterprises, the monthly changes in responses may be due more to seasonal variations than to current economic conditions. But the balances of opinion, which serve as summary indicators of the responses, are designed to track monthly changes in current economic conditions in industry. We therefore need to remove the seasonal information contained in the series obtained from the aggregation of individual responses, so as to isolate the information pertaining to the economic cycle. That is the purpose of seasonal adjustment.

2.5.1. - General method

Selected balances of opinion are seasonally adjusted once a year[5]. The operation is performed with *X*11-*ARIMA* software for monthly and quarterly series. To produce a seasonally adjusted series from a raw series, we subtract the seasonal-coefficient series from the raw series:

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

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

---

[5] The percentages of "increasing", "unchanged", and "decreasing" responses are not seasonally adjusted; only the balances of opinion are.

The seasonal-coefficient series are updated each year to reflect the influence of the latest data points on seasonal-coefficient values. The updates thus require an annual recalculation of the seasonally adjusted series.

All published questions are seasonally adjusted. Zero-value seasonal coefficients denote series that have been seasonally adjusted but display finally no seasonality.

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*:

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

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 slightly over a year apart.

2.6. - Storage of aggregated data

Raw series, seasonally adjusted series, results from the constant sample, and results from the current sample are stored.