
***PART TWO:
SURVEY EXECUTION***

PART TWO: SURVEY EXECUTION

2.1. - Survey sample

The sample of the Quarterly Survey of Industrial Investment is similar to that of the Monthly Business Survey on Industry (*see* Biau and Bonnefoy, 2007).

2.1.1. - Sampling frame and sample stratification

The unit surveyed in the Investment Survey is the enterprise. The Survey uses a sector-based approach. The sampling frame is the latest available Annual Enterprise Survey (EAE).

INSEE uses two criteria to stratify the population:

- The activity sector defined in the Nomenclature Économique de Synthèse (NES) at level 36 with two subsets for the food industries (meat and dairy industries; other food industries). The sector for each enterprise is defined by its principal activity under level 700 of NAF Rév. 1.
- The number of employees, as indicator of enterprise size. Size categories are as follows: category 1: fewer than 100 employees - category 2: 100-499 employees - category 3: 500+ employees.

As in the Monthly Business Survey on Industry, all of the largest enterprises are surveyed. They consist of enterprises with 500+ employees or turnover exceeding €150 million. These units form the “exhaustive” stratum. For the other strata, we choose the sample by means of stratified sampling without replacement, using an allocation proportional to turnover.

The sample—which is an enterprise panel—is maintained from one Survey to the next. However, the sample is partially replaced at regular intervals owing to (1) economic developments such as deaths and restructuring of enterprises and (2) the need to preserve a sample of broadly constant size, and offering relevant coverage of the industrial sector.

2.1.2. - Sample replacement

The Investment Survey sample is thus regularly updated from the sample used for the Monthly Business Survey on Industry. This updating mainly involves “cleaning” followed by additions to the sample.

“Cleaning” consists in (1) removing from the sample the units that have left the Survey’s scope of coverage (enterprises with fewer than 20 employees or with a non-industrial principal activity) and (2) incorporating the units that have been added to the “exhaustive” stratum since the previous replacement.

We enlarge the sample by adding enterprises taken from the Annual Enterprise Survey to the “non-exhaustive” strata. Using a stratified and balanced random sampling procedure with the aid of the SAS CUBE⁷ macro, we ultimately obtain a sample of approximately 4,000 enterprises.

⁷ Available on the INSEE website at www.insee.fr/fr/nom_def_met/tools_stat/cube/documentation_cube_web.pdf.

2.2. - Collecting information

2.2.1. - Data collection

We send the questionnaires to the sampled enterprises at the very end of the month prior to the Survey month. The responses must reach the Institute—which keys the questionnaire data—by around the 10th of the Survey month.

In each wave, INSEE issues an initial reminder by sending follow-up questionnaires by postal mail to firms that have not responded by the return date printed on the initial questionnaires.

Around the 20th of the Survey month, INSEE sends a new reminder by telephone or email to the largest non-respondents. This follow-up generates further improvement in the Survey response rate and therefore in the quality of results.

2.2.2. - Post-collection checks

INSEE's Business Surveys Division keys the data as the questionnaires arrive. A system of *checks* alerts survey clerks to “suspect” responses, thereby limiting keying errors.

Responses are not necessarily consistent within a single wave or from one wave to another. A preliminary procedure, called *pre-extraction (pré-dépouillement)*, therefore aims to correct some aberrant values in the individual-data file by checking responses for consistency and imputing values in the simplest cases.

We are particularly attentive to annual investment expenditures. First, we make sure that the amounts are consistent with enterprise size, measured both by the number of employees and by turnover. Second, the consecutive estimates of investment for a given year must be all the closer from one wave to the next as the waves cover past years.

We also perform several other types of checks:

- on consistency of “structural” data: large annual variations in turnover and workforce size are flagged, as they may signal errors in the enterprise's responses or in response keying
- on coding allowed for keying responses to questions about business conditions: the coding must comply with the options provided in the questionnaire⁸
- on classifications: by introducing a table of classifications of activities in the keying application, we can check the validity of the keyed codes.

When our checks indicate a high risk of error on an enterprise's responses, we contact our correspondent in the enterprise by telephone. We devote special attention to large firms. Depending on the information provided by the correspondent, we either validate or correct the responses initially regarded as “suspect”.

At the end of the checking process, we therefore obtain a clean data base that is used for the final extraction of Survey data.

⁸ Appendix 5 gives the main questions, as well as the authorised responses and coding.

2.3. - Processing the Survey data

After keying, we adjust questionnaires for non-response (*see* §2.3.1) and aggregate the information into summary indicators (*see* §2.3.2 and §2.3.3).

2.3.1. - Treatment of non-response

We perform two operations on our 4,000-enterprise sample to make the results representative of the total industrial sector. The first operation addresses partial non-response, i.e., when the enterprise has returned an incomplete questionnaire. Whenever possible, we impute values for this purpose. The second operation concerns total non-response. In this case, we reweight the raw results to ensure consistency with the structure of the industrial sector.

2.3.1.1. - Construction of a “completed” sample

Generally speaking, imputation or editing for non-response aims to improve the quality of results in terms of bias and/or variance. The method typically adopted in business surveys is to copy an earlier response.

The data used in this type of operation are not always comparable. They may concern different periods or, when they apply to the same variable, they may have been estimated at different dates. In any event, the statistical objects handled are visibly distinct.

There is a trade-off between the potential gain provided by an additional datum and the risk entailed in using a potentially erroneous or obsolete datum. In high-frequency surveys such as our Monthly Business Survey on Industry, the net outcome is regarded as broadly positive.

The Investment Survey, however, is somewhat more delicate to handle. For highly scattered variables such as investment, or highly inert variables such as turnover, even an old or rough figure is clearly of great value and improves the quality of the estimate. In such cases, response completion seems fully justified.

By contrast, for short-term variables such as the six-month change in investment, the range of possible values is narrower, but they change rapidly and are often revised to a sometimes significant extent. Replicating data that are too old may therefore entail substantial errors. The problem is that most questions on business conditions in the Investment Survey are asked every six months; some are asked only once a year.

We therefore examined the relevance of response completion on a case-by-case basis for all Investment Survey questions.

Imputation for non-response or missing values is restricted to turnover, workforce size, annual investment, and variables on investment purposes. We use the “cold-deck” method, i.e., a value that is not available will be replaced by the equivalent value (i.e., pertaining to the same year) most recently reported (or inferred) in one of the two previous surveys.

For turnover or workforce size in year N, the completion algorithm chooses the most recently reported value for year N or, in its absence, the value retrieved from an external source such as the business register (SIRENE), accounting data from commercial-court registries, etc. Otherwise, the algorithm chooses the value declared for year N-1.

For total annual investment in year N, we use different completion procedures according to intended use:

- To determine annual investment growth, the quality of the results depends heavily on the consistency (in terms of scope, definition, response date, and so on) between the figures reported for the three years. In this case, we perform completion only when none of the three annual amounts is provided. Also, we do not carry forward earlier responses unless they do not date back beyond two waves (in other words, if they are no more than six months old). The few other problematic situations do not warrant the development of more complex completion procedures.

- For other survey questions, annual investment may serve as a weighting coefficient (*see* §2.3.2 and following). In this case, consistency between successive years is not crucial, and the best option is to include as much information as available. Accordingly, we copy the previous figure reported for the same year.

For the “Purpose of investment in France” box included in the April and October waves, if the enterprise has not answered any of the questions in the box, we copy the response given in the Survey for the previous six-month period.

We do not perform the completion procedure for non-response for all the other Survey variables.

The completion procedure does not concern the previous wave, whose results we therefore do not reprocess.

2.3.1.2. - Weighting the Survey results

We use two weighting systems to aggregate results in a manner that will represent the actual structure of industry as faithfully as possible:

- “Primary” weightings serve to aggregate results in each primary stratum.
- “Secondary” weightings serve to aggregate different strata together.

By and large, the strata chosen to determine these weightings are the same as those used to select the sample, i.e., a combination of the enterprise’s activity sector (NES at 36-heading level) and size (three categories).

Depending on the type of question, primary weightings reflect the share of the unit’s turnover or reported investment in total stratum turnover or investment. Secondary weightings are external variables computed, for each stratum, from either (1) combined turnover in year N-2, to tabulate responses on scrapping (*see* Appendix 5) or (3) combined investment in year N-1 to tabulate responses to other questions.

We recalculate secondary weightings in every October wave from the latest Annual Enterprise Survey (EAE) available for year N-2⁹. They remain constant until the July wave in the following year.

Combined turnover for year N-2 is taken directly from the EAES. Combined investment for year N-1 is estimated: for each stratum, we take the investment growth rate between years N-2 and N-1 estimated in the latest Investment Survey wave and apply it to the EAE figure for combined investment in year N-2.

2.3.2. - Calculating annual investment growth rates

Enterprises are asked to report their annual investment every quarter. These questions are central to the Survey, as they allow us to estimate expected and actual annual variations in industrial investment. The results are the most informative of the Survey, and the ones most widely reported in the media.

2.3.2.1. - Choosing a method that is robust to atypical changes

However, because of the quantitative nature of these questions and the diversity of individual investment behaviours, we need a special processing method. The most logical choice to estimate aggregate changes from individual data would appear to be the ratio estimator. But it lacks robustness, for an enterprise with atypical behaviour can cause a wide swing in the aggregate result. It is therefore essential to apply a method that will be robust to atypical changes.

There are two methods available: the “Large Investors” (LI) method and the more theoretical method called “GM-estimators”, described by Ravalet (1996). For published Survey results, we use only the first method: given its resemblance to the ratio method, it displays the robustness of complex estimators such as GM-estimators, while enabling results to be easily understood.

⁹ They were taken from the FUTE (Fichier Unifié Total Entreprise) file until 2006.

2.3.2.2. - Principle of “Large Investors” (LI) method

The LI method consists in underweighting certain enterprises that are judged to be non-representative of the total sector examined. The method accordingly distinguishes between two types of enterprises:

- “Extrapolatable” enterprises, which, as their name indicates, are representative of the stratum. The growth rate of their investment can thus be extrapolated to the entire stratum, with the exception of “Large Investors”, if any.
- “Large Investors” (LIs), which are regarded as non-representative of the total stratum. In other words, they represent only themselves.

An enterprise is classified as an LI if it is viewed as atypical or influential:

- An enterprise is viewed as “atypical” when its investment varies too sharply between two consecutive years (a tenfold or greater increase or decrease).
- An enterprise is viewed as “influential” when it contributes heavily to the results aggregated at NES 16-heading level, i.e., the initial publication level. The classification of enterprises as “influential” enterprises is performed manually. It is, however, guided by “change indicators” that report the difference in the growth rate when the enterprise status is changed from extrapolatable to “Large Investor”. When the change indicator exceeds a given threshold, the enterprise is regarded as influential. The threshold, which is sector-specific, is always less than one point: in other words, no extrapolatable enterprise contributes more than one point of growth at NES 16 level.

2.3.2.3. - Calculation of growth rate for a given stratum

We thus estimate the growth rate for the stratum as the weighted sum of investment growth rates for “Large Investors” and extrapolatable enterprises. More specifically, the investment growth rate between years N-1 and N for a given stratum s, noted $t_{s, n-1 \rightarrow n}$, is estimated as follows:

$$t_{s, N-1 \rightarrow N} = \frac{i_{s, N-1}^{GI}}{I_{s, N-1}} t_{s, N-1 \rightarrow N}^{GI} + \left(1 - \frac{i_{s, N-1}^{GI}}{I_{s, N-1}} \right) t_{s, N-1 \rightarrow N}^{extrapolatable}$$

with: $t_{s, N-1 \rightarrow N}^{GI}$:	Investment growth rate of stratum s between years N-1 and N for “Large Investors” in the sample.
$t_{s, N-1 \rightarrow N}^{extrapolatable}$:	Investment growth rate of stratum s between years N-1 and N for extrapolatable enterprises in the sample.
$I_{s, N-1}$:	Total investment by population of stratum s in year N-1. These totals determine the secondary weightings.
$i_{s, N-1}^{GI}$:	Total investment by “Large Investors” in stratum s in year N-1.

2.3.2.4. - Calculation of growth rate for a group of strata

Some primary strata prove to be too small for us to apply the “Large Investors” method. We therefore group them together for the sole purpose of calculating investment growth rates (*see* Table 5).

Table 5:
Grouping of strata in January 2007 Survey wave

Sector (NES 36)	Size-category groupings
B1 Meat and dairy industries	1 and 2
B2 Other food industries	1 and 2
C1 Manufacture of clothing articles and leather products	2 and 3
C2 Publishing, printing, reproduction	1 and 2
C3 Pharmaceuticals, perfumes, and soap and cleaning preparations	1 and 2
C4 Manufacture of domestic equipment	1 and 2
D0 Manufacture of motor vehicles	1 and 2
E1 Building of ships and boats, manufacture of aeronautical equipment, manufacture of railway locomotives, rolling stock	1 and 2
E2 Manufacture of metal products, machinery, and equipment	1 and 2
E3 Manufacture of electric and electronic equipment	1 and 2
F1 Mining and quarrying except energy producing materials, manufacturing of other non-metallic mineral products	1 and 2
F2 Manufacture of textiles	2 and 3
F3 Manufacture of wood, wood products, pulp, paper and paper products	1 and 2
F4 Manufacture of chemicals, rubber, plastic and chemical products	1 and 2
F5 Manufacture of basic metals and fabricated metal products	1 and 2
F6 Manufacture of electric and electronic components	1, 2 and 3
G1 Extraction of coal, crude petroleum, gas and uranium; manufacture of coke, refined petroleum products, and nuclear fuel	1, 2 and 3

We then aggregate investment growth rates for strata s between years $N-1$ and N by applying adjustment coefficients, called “secondary weightings”, that reflect each stratum’s investment share in the actual population. For a set of strata H , the coefficient is equal to:

$$t_{H, N-1 \rightarrow N} = \sum_{s \in H} \frac{I_{s, N-1}}{\sum_{s \in H} I_{s, N-1}} t_{s, N-1 \rightarrow N}$$

We calculate the growth rate between the first two years covered by the Survey questionnaire by directly applying the secondary weightings (computed from total investment in year $N-1$, *see* §2.3.1). The growth rate between years N and $N+1$ is estimated from updated secondary weightings (computed from total investment values for year N obtained by applying the previously determined growth rates to the values for year $N-1$).

2.3.3. - Processing of other short-term questions

Amongst the other Investment Survey questions, the largest group consists of ordinal-scale qualitative questions. They are processed in accordance with a general principle described below.

2.3.3.1. - Calculation principles for ordinal-scale questions

Like the growth-rate questions, ordinal-scale questions require primary *intra*-stratum aggregation and secondary *inter*-strata aggregation.

The primary aggregation of ordinal-scale qualitative questions with two, three or five choices (*see* Table 6) consists in calculating, for each primary stratum s , the weighted proportions of individual responses to each choice. The procedure is performed only on the sub-sample of enterprises that have been weighted and have responded to the question.

Table 6:
Description of ordinal-choice qualitative questions

Two-choice questions	Three-choice questions	Five-choice questions
y = 1 ⇔ yes	y = 1 ⇔ increase	y = 1 ⇔ very stimulative
y = 2 ⇔ no	y = 3 ⇔ stable	y = 2 ⇔ stimulative
y = . ⇔ no response	y = 5 ⇔ decrease	y = 3 ⇔ no influence
	y = . ⇔ no response	y = 4 ⇔ restrictive
		y = 5 ⇔ very restrictive
		y = . ⇔ no response

The primary weighting, p_i , is determined either from turnover or from one of the three investment figures. Primary aggregation consists of the following:

- for a two-choice question: the weighted percentage of enterprises responding “yes” among those having responded “yes” or “no”¹⁰ in stratum s , i.e.:

$$Y_s^1 = 100 \times \frac{\sum_{i \in s} p_i \times 1_{y_i=1}}{\sum_{i \in s} p_i} \quad \text{where } 1_{y_i=1} \text{ equals 1 when } y_i=1 \text{ and 0 if not;}$$

- for a three-choice question: the difference between the weighted percentage of enterprises expressing a positive opinion (“1”) of the situation and the weighted percentage of enterprises expressing a negative opinion (“5”) in stratum s , i.e.: $Sol_s = Y_s^1 - Y_s^5$;
- for a five-choice question: the difference between the weighted percentage of enterprises expressing a positive opinion (“1” or “2”) of the situation and the weighted percentage of enterprises expressing a negative opinion (“4” or “5”) in stratum s , i.e.: $Sol_s = Y_s^1 + Y_s^2 - Y_s^4 - Y_s^5$.

The indicator at aggregated-strata level consists of the weighted average of primary stratum indicators (Y_s^1 or Sol_s). Thanks to the weighting, we can allow for the relative weight of the stratum in the Survey’s total scope of coverage. The secondary weightings are exogenous (*see* §2.3.1). They are noted P_s and are determined from either stratum investment or stratum turnover, depending on the question. The secondary aggregation for a set H is:

$$Y_H^1 = 100 \times \frac{\sum_{s \in H} P_s \times Y_s^1}{\sum_{s \in H} P_s}$$

2.3.3.2. - Six-month change in investment

The two qualitative questions (increase, stable, decrease) concern the past and expected six-month change in investment (*see* §1.2.3 and Appendix 2). They have been asked in April and October since 1991 and in January and July since 2003.

Both questions are processed using the general method for three-choice questions described above. We use the current sample, without completion. The reason is that horizon differences make it impossible to define a consistent completion rule for all waves: the questions are asked quarterly but cover two consecutive six-month periods. By contrast, the primary weighting is based on the completed investment figure.

These are the only sub-annual questions in the Survey. They display high seasonality, which requires adjustment. Seasonal adjustment consists in subtracting from the series the half-difference between (1) the average of unadjusted results for the first six-month periods and (2) the average of unadjusted results for the second six-month periods.

¹⁰ Enterprises that have not answered the question are not included in the calculation of the balance.

2.3.3.3. - Investment purposes

The first question set deals with investment purposes. We ask respondents to classify past-year and current-year investment under the following five headings:

- replacement of used equipment, upkeep, maintenance
- modernisation and streamlining
- increase in productive capacity for existing products
- introduction of new products
- other purposes: safety, environment, working conditions, etc.

Each of these headings is associated with a dummy variable and a percentage showing the approximate share of investment completed (or planned) chiefly for the listed purpose. We perform checks and edits to make these two information items consistent. For example, if the respondent has provided a percentage figure, the dummy will necessarily be set to “1”. Responses also undergo a completion procedure: if the respondent has not specified the investment purpose, we copy the corresponding response provided in the Survey for the previous six-month period.

We transform the dummies into breakdown variables in which each choice is weighted $1/R$, with R the number of choices selected by the enterprise. In each primary stratum, the result is determined as follows: once normalised, the shares of each choice are weighted by the investment figures reported by the enterprise in the same wave and for the year examined. We then aggregate the stratum results, adjusting for the relative investment shares in the corresponding year¹¹.

The second question block allows respondents to select the main purpose of modernisation investment among the following three choices:

- automation of existing manufacturing processes
- introduction of new manufacturing methods
- energy savings.

As with the variables for main investment purpose, we convert this variable into a proportion, in which each choice is weighted $1/R'$, with R' the number of choices selected by the enterprise. Response completion and tabulation follow the same principle as the previous questions.

2.3.3.4. - Productive capacity

In the April wave, industrial firms are asked to give the past and expected annual change in their productive capacity (see §1.2.4). The data collected are both qualitative (Increase / Stable / Decrease) and quantitative, for the enterprise also provides the order of magnitude of the change (for stability, the quantitative variable is set to 0).

These questions are tabulated on a current-sample basis and weighted by the enterprise's share of total turnover in its primary stratum. The secondary weighting, applied to the stratum, is determined from the primary-strata turnover values calculated for the Monthly Business Survey on Industry.

In the past, the quality of the results obtained for the quantitative question was found to be poor. The gap between forecast and outcome was fairly significant. Moreover, the results seemed to overestimate the change in productive capacity estimated from national-accounting data and from the capacity utilisation rate indicated by the Monthly Business Survey on Industry. For this reason, we now publish only the balances of opinion.

A second question block concerns equipment scrapping. Respondents are asked to give the main reasons and to qualify the magnitude of the change. An initial two-choice variable (yes/no) tells us whether any equipment was scrapped in the year examined. If the answer is “yes”, four reasons are offered:

- wear and tear, obsolescence

¹¹ Year N-2 in the January, April, and July waves of year N; year N-1 in the October wave of year N.

- installation of more efficient equipment
- shut-down of capacity for old products
- other.

Two three-choice variables define the change in scrapping (1) between years N-2 and N-1 and (2) between years N-1 and N.

We process these questions on a current-sample basis and weight them by the corresponding non-completed investment figure. The two-choice and three-choice variables are tabulated in accordance with the same principle as the previous questions. We use a special procedure to process the scrapping-purpose variable.

2.3.3.5. - Factors influencing investment decisions

Until 1991, industrial firms were asked to list obstacles to investment, such as hiring problems that made investment pointless, heavy indebtedness, supplier-related problems, and inadequate cash flow. These questions were later dropped. Conversely, at the European Union's request, the Survey has incorporated questions about the influence of various factors on investment since 1987.

Respondents are asked to rate the influence of nine possible factors over two consecutive years. The list includes "domestic-demand outlook", "cash flow", "interest rates", etc., on a five-level scale ranging from "very stimulative" to "very restrictive". We process these variables without completion, using the method applied to a five-choice variable (*see* §2.3.3.1). The primary weighting is the corresponding, non-completed investment figure.

