Study on

Sample Design for Industry Surveys

ECFIN/2003/A3-03

Final Report

November 2004

Laurent Donzé, Richard Etter, Nora Sydow, Oliver Zellweger

Swiss Federal Institute of Technology (ETH)
Swiss Institute for Business Cycle Research (KOF)

ETH Zentrum WEH
CH-8092 Zürich, Switzerland

CONTENTS

FIGURES

TABLES

1 Introduction

Business tendency surveys (BTS) are qualitative economic surveys. Their use is mainly for short-term economic analysis, including the prediction of turning points in the economic cycle. BTS are a complement of quantitative statistics but are usually conducted on a higher frequency, and, in addition, the results are published faster. The survey asks the responding company about the current business situation and short-term developments within the EU. The surveys are carried out on a national level but are co-ordinated and aggregated by DG ECFIN of the European Commission.

The survey can fulfil the aim mentioned only if it possesses the main properties necessary for any statistics. As noted by Eurostat[1], Statistics Finland, or Statistic Canada, the quality of statistics can be defined with respect to the following criteria:

1. Relevance of statistical information;
2. Correctness and accuracy of data;
3. Timeliness and promptness, punctuality, or delivery on schedule;
4. Accessibility and transparency / clarity of data;
5. Comparability of statistics;
6. Coherence and consistency / uniformity of data;
7. Documentation.

Point 1 (relevance of statistical information) is undisputed whenever high quality of BTS data is given. Point 3 of this list has been optimised in several steps during the past years. The introduction of the new statistical classification of economic activities (NACE rev 1) has increased the comparability (point 5) of BTS with other statistics considerably. Moreover, the core questions are now almost identically formulated throughout all EU countries and the periodicity is the same. Point 4, 6 and 7 are important aspects of the quality of a statistics yet within this project we will concentrate on parts of the correctness and accuracy of the data: on the sample design of the survey. Accuracy is defined as the closeness between the estimated value and the (unknown) true population value.

On a national level correctness and accuracy of the collected data depends on clear and precise questions, on an optimal sample design and effective follow-up activities.

[1] These attributes, recommended by the European Expert Leadership Group on Quality are outlined in Eurostat (2002), Definition of quality in statistics, working paper for the 5th meeting of the Working Group of Assessment of the quality in statistics, 2-3 May 2002.

But for the aggregation of the data to results for the European Union as a whole, the requirements are higher. For higher comparability of the results, questions, sampling procedures, weighting schemes and imputation techniques should be harmonised.

Until now, the sample design of the BTS in the EU differs from country to country. The decentralised mode of sample design has the advantage of allowing institutes to take, in a best possible way, country-specific conditions into account. On the other hand, a multitude of sampling schemes can affect transparency and the international comparability. Moreover, it could lead to quality variations across countries as different sampling and non-sampling errors may affect the survey results.

The purpose of this study is to make general recommendations concerning “best practices” in conducting business tendency surveys; it is focussed on the issue of sample design for surveys in manufacturing. In the frame of the joint EU programme of business and consumer surveys sample design is an area of discretion for the individual collaborating institutes. Institutes were, however, asked to give adequate information about various elements of the sample design.

The paper is structured in four parts. Section 2 compares the different elements of the national survey designs collected by a questionnaire. The questions were divided into several subgroups: Questions on the frame list, the sampling methods and the weighting procedures were key elements. Information on the response rate, handling of missing data and the quality assurance framework completed the questionnaire. Section 3 gives some hints to the design effects in BTS data of the different countries. In section 4, six countries are selected to analyse their sample design in-depth. Individual questionnaires were adapted to the country-specific situation. An individual description and assessment is given. The study finishes with a technical report including recommendations.

2 Sample design in the EU15 countries

2.1 General remarks

It is obvious that a harmonised sample design within the EU countries would improve the comparability of the country results. Moreover, the aggregation of the harmonised survey results of the countries to total EU increases the transparency of the underlying aggregation process and thus helps to interpret the EU results accurately. Up to now a contrasting survey of the sample design of the BTS conducting institutions has been missing. This report is intended to fill this gap.

Relevant information was collected by a postal survey. The questionnaire and a covering letter were sent to the managing directors of the institutes of the 15 countries that made up the EU before enlargement (EU15) and Poland, and also to the heads of the unit in charge of conducting the industrial trend surveys. By the end of the deadline for submissions, we received fifteen questionnaires, the remaining institute was contacted in a follow-up by a reminding letter and a reminding e-mail, without success. In general, the responding institutes were collaborating and ready to detail their answers.

The questions were asked in a general manner and respondents were given the chance to specify their answers. The main purpose of our survey was to get a picture of the common practice applied in the different EU countries; a overview of answers is shown in table 1.

In the following subsections the different parts of the questionnaire and the results are presented using some figures and comments. The sequence of the ensuing subsections are similar to the structure of the questionnaire. After a short introduction to each topic an evaluation of the answers is provided. A copy of our questionnaire can be found in the appendix of this report.

table 1: Synoptic summary

countries	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Sources of frame list															
business register	x			x	x	x		x		x	x		x	x	
government register	x	x					x		x			x			
own sources of information	x		x							x			x		
company lists	x		x							x					
other										x					x
Frame list															
frame list includes all companies	x	x							x						
Cut-off															
branches			x	x											
size			x		x	x	x				x	x	x	x	x
regions															
other								x		x					
surveyed unit = sampled unit		x							x		x	x			x
Sampling method															
simple random sampling										x					
stratified sampling with simple random sampling	x	x					x	x	x			x	x		
sampling with probability proportional to size						x									
cluster sampling with simple random sampling															
exhaustive survey with cut-off					x									x	
other			x	x							x		x		x
Response rate															
mode of survey															
interview												x			
self completion postal survey	x	x	x	x	x	x	x	x	x	x	x		x	x	x
e-mail survey		x		x								x			
web survey					x								x	x	x
follow-up activities	x	x	x	x	x	x	x	x	x	x	x		x	x	x
average response rate after follow-up 2003	0.75	0.75	0.85	0.95	0.80	0.92	0.61	0.38	0.62	0.50	0.55		0.82	0.99	0.55
award scheme			x	x	x	x	x	x		x	x	x	x		x
Missing data															
explore nonresponse behaviour			x				x								
deal with item non response		x	x			x	x		x	x		x			
deal with unit nonresponse		x	x			x	x	x	x	x	x	x			
weighting procedures									x						
auxiliary information for weighting															
firm size (number of employees, turnover)	x	x	x	x	x	x	x	x	x	x		x	x	x	
sector			x	x			x	x	x	x					
branches			x	x		x	x	x				x	x	x	
other	x						x				x	x	x		
Quality assurance															
comparison with benschmark series	x		x	x			x	x	x	x			x		x
sensitivity analysis to examine weighting factors							x			x					

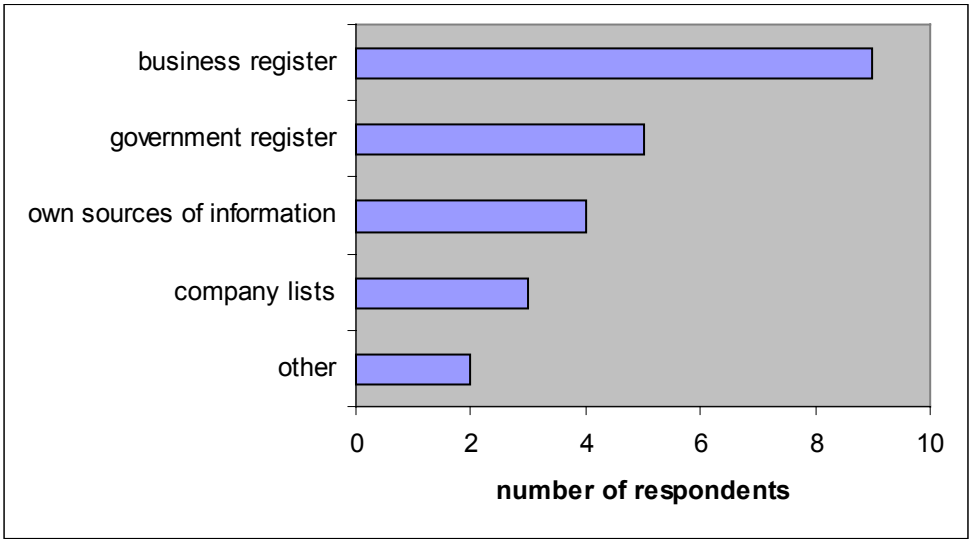
We disseminate only anonymised data

2.2 Sample frame

BTS are conducted among firms. Demographic structure variables, addresses and economic activity of a firm have to be known to build up a sample framework. But where does the institutes get their data from? A government register, at its best, collects data about each existing firm in a country and displays the complete survey population (the manufacturing sector, in our case). As it will be shown later, not all institutes are using this kind of register data as the sample frame.

As can be seen in figure 1, the majority of the survey conducting institutes uses information from an administrative source such as business registers or any other type of governmental register. 80% (12) do not cover the whole manufacturing industry mainly because of cut-off limits according to some size measures (see figure 2).[2]

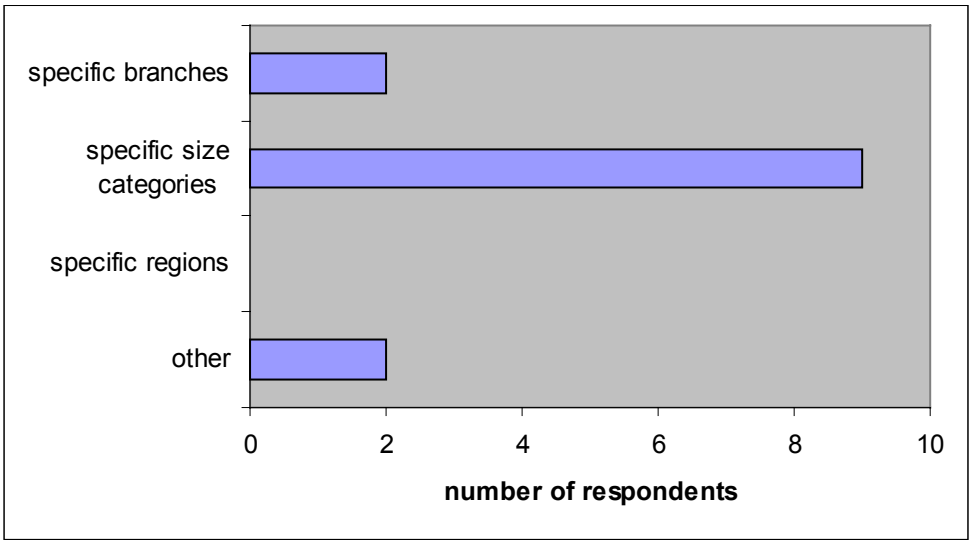
figure 1: sources of information entering the frame list



source: question 1.1; several answers possible

[2] Please note: the total number of institutes is mentioned in brackets.

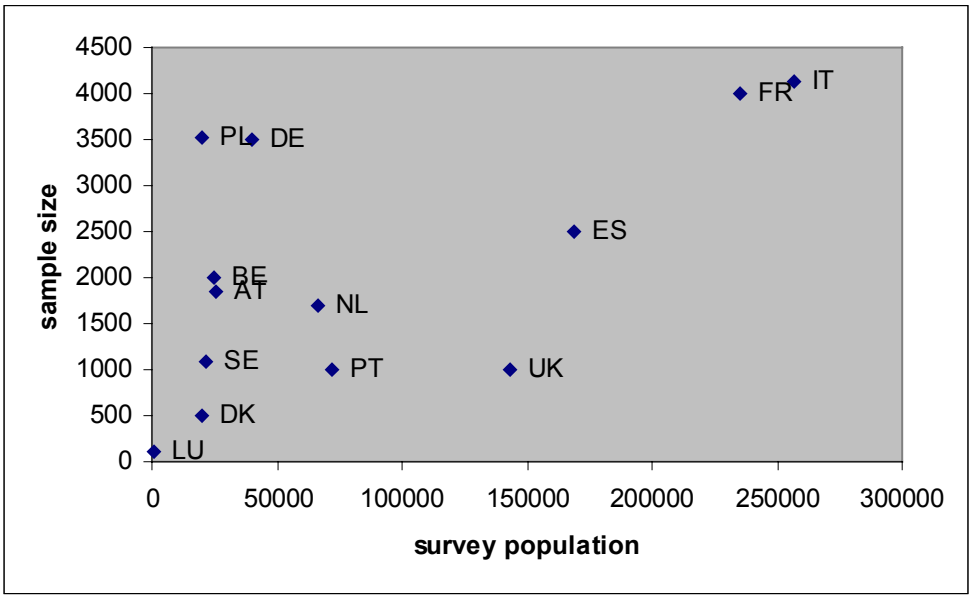
figure 2: units missing in the frame list



source: question 1.3; several answers possible

Taking the reported data on the actual sample size and the population size (figure 3) in the manufacturing survey leads directly to the sampling fraction.

figure 3: sample size and survey population by country



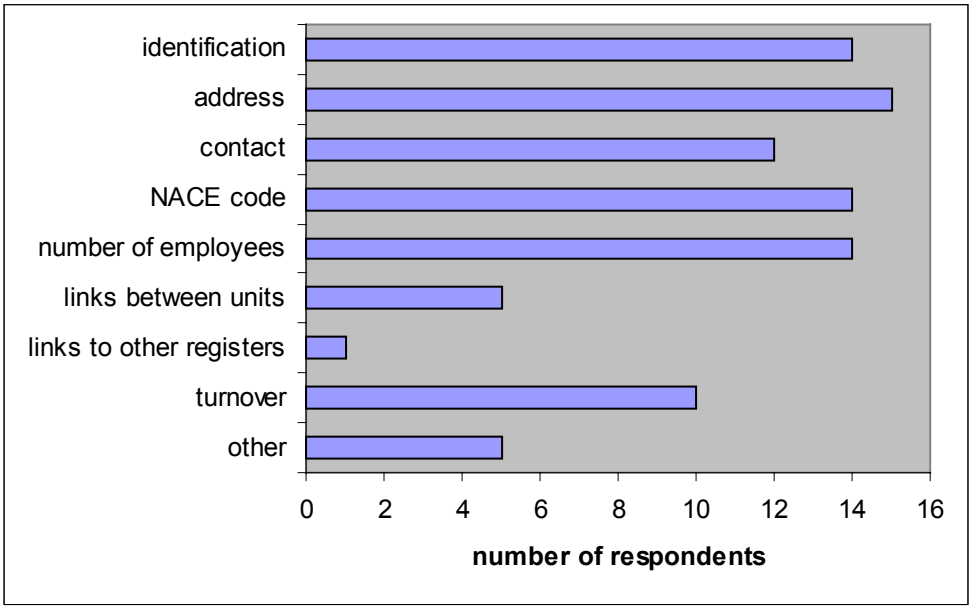
source: question 1.2 and 3.6; country codes: BE Belgium, DK Denmark, DE Germany, ES Spain, FR France, IT Italy, LU Luxembourg, NL Netherlands, AT Austria, PL Poland, PT Portugal, SE Sweden, UK United Kingdom. Data for both Finland and Greece are missing.

It is tempting to speculate that the sampling fraction is inversely related to the size of the country. This proposition becomes intuitively clear as the sample size depends mainly on precision considerations of estimates and not directly on the population size. Therefore the sampling fraction is determined to a great deal by the sample

size. As the optimal sample size is by definition smaller than the population size, the sampling fraction decreases gradually with the size of the survey population.

The data contained in the frame list considers different characteristics which are helpful to identify the units. The demographic structure is assured by providing the NACE code and the number of employees for all surveying institutes (see figure 4). The availability of the addresses ensures the delivery of the questionnaire to the right entity. Cross-linking between different units would be helpful to avert duplication. This kind of information, however, is not contained in all frame lists (67%). It is quite astonishing that 67% (10) of the frame lists even contain such sensitive data as value added and/or turnover of a company. All institutes update characteristics and thus the frame regularly. A cycle of 1-2 years enables a constant quality of the frame list.

figure 4: characteristics contained in frame list



source: question 1.4; several answers possible

Nevertheless 80% (12) of the institutes are applying some procedures to maintain or improve the quality of the frame list besides these updates. These procedures consist in eliminating duplicates (11) and in adapting characteristics (12).

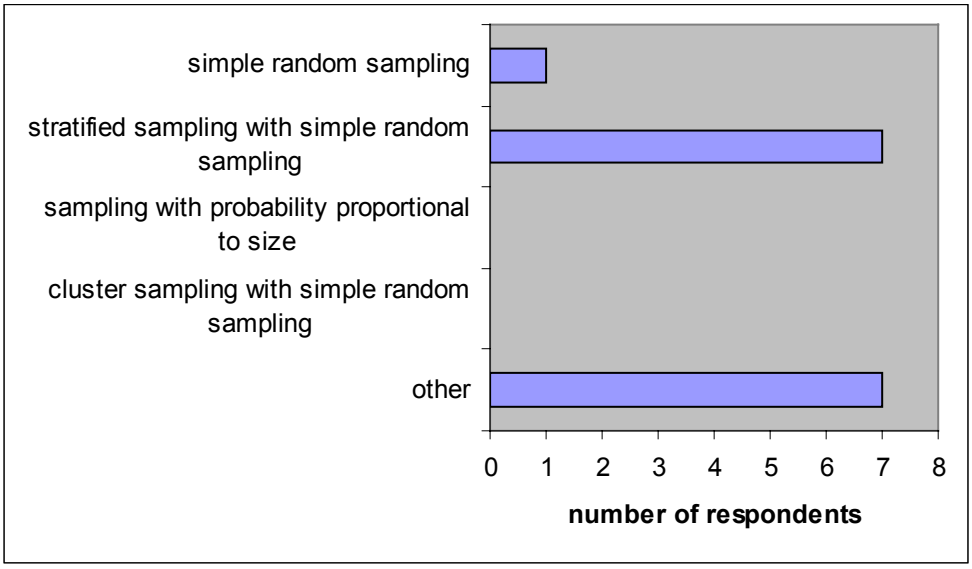
In some cases, the legal entity is not the appropriate level to receive the most adequate information of the economic development. Different production units of a company can evolve differently. One measure is to provide several questionnaires to a company for each kind-of-activity unit although the concept of the unit of the frame list is contradicted and the aggregation of the results may lead to distortions. Two thirds (10) of the institutes deliver more than one questionnaire if necessary.

2.3 Sampling methods

A sample is used when surveying the whole population would be too laborious and/or expensive. A sample is drawn in order to represent the structure of the population in a best way. To keep costs low the sample is kept as small as possible. Under certain conditions, statistical inference allows the formation of conclusions about the parameters from a sample taken from a larger population. In BTS a selected sample is kept for a longer period and not redrawn monthly and thus resembles a panel. On the one hand, this constancy will lead to an increasing response rate, but on the other hand, the constant sample may drift away from the true structure of the industrial sector.

To cope with the heterogeneity of the survey population, most of the institutes use stratified sampling with simple random sampling. The answers given to question 2.1 are summarised in figure 5. Among the answer category “other” one institute reported applying quota stratified sampling, the second one voluntary stratified sampling with some exhaustive strata. The stratification criteria are size and branch (5), gross value added (1), branch (1) and turnover (1). The remaining five institutes ticking “other” are split in those surveying all units on the frame list, the largest companies and one contacts business organisations. The last two did not specify their sampling method.

figure 5: sampling methods

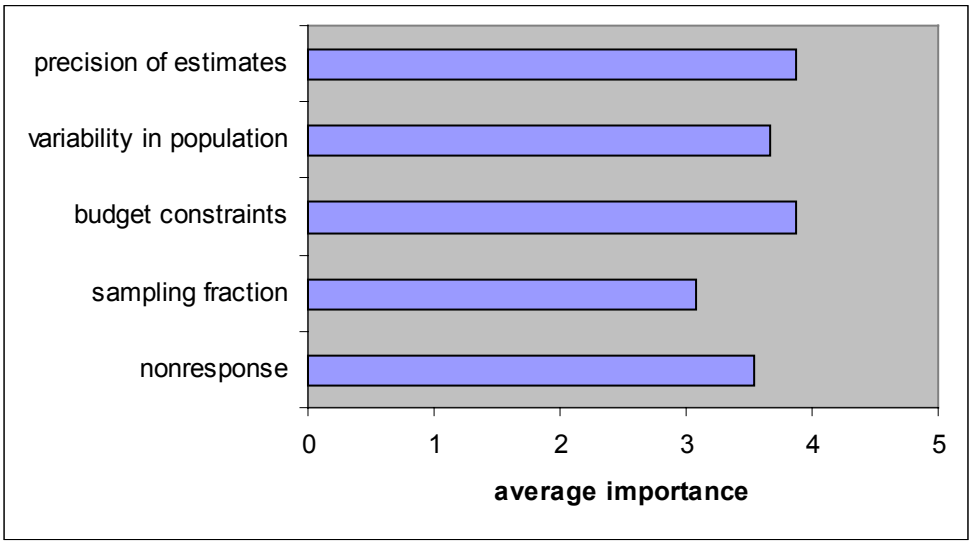


source: question 2.1; single answer

Only one of the institutes uses some sort of rotation scheme in order to reduce the reporting burden; but some of the remaining have already taken this into account by choosing a cut-off for specific size categories.

Figure 6 shows that on a whole all key factors except sampling fraction are important when deciding on sample size. The results also point to the implicit trade-off between maximising the precision of estimates and the underlying costs. Most countries reporting response rates above 75% rated the precision of estimates very highly. There is an inverse relation between budget constraints and the response rate of a survey. This result supports the intuitive fact that more financial resources may improve the quality of a survey at least to some extent.

figure 6: key factors for sample size



source: question 2.4; several answers possible

2.4 Response rate

The panel character of BTS data leads to a higher response rate compared to a recurring survey conducted with a sample drawn for each survey wave. Nevertheless the response rate has to be kept continually on a high level to assure comparability over time.

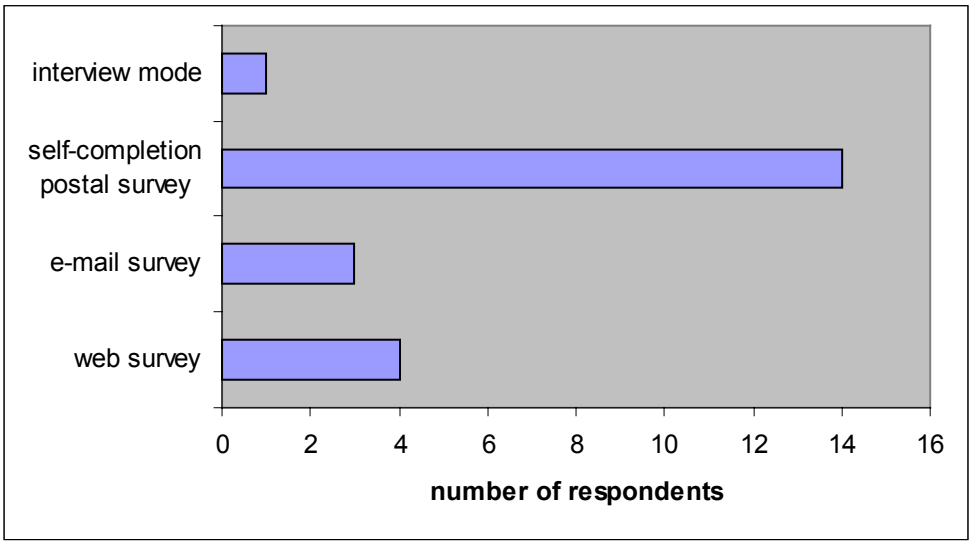
In principle, one could think of making participation compulsory to increase survey return. Firms in Luxembourg, France, Portugal, Italy, the Netherlands and Spain are in fact required to take part in the survey on manufacturing by legal obligation. Yet response rates in these countries were not affected markedly by this legal requirement, that is, comparing response rates with countries without legal obligation did not show any substantial differences. According to our questionnaire, to be fair, we have no information on the type of imposed sanctions or other measures to enforce regulation. However, a compulsory survey is likely to affect the quality in the

collected data negatively. Therefore, if there is no impact on the response rate, legal obligation could be scrapped.

Advertising can also help to keep the number of respondents constant. 60% of our respondents have reported to use this method in order to maintain a constant sample size. In addition to advertisements on the Internet, some institutes are contacting business sector associations in order to keep their sample units participating or to acquire new firms.

The mode how the questionnaire is addressed and delivered to the firms may also influence the response rate. It is important to provide the questionnaire in the most convenient mode to the recipient in order to maintain or increase response rates (figure 7).

figure 7: survey mode



source: question 3.3; several answers possible

Almost all institutes (14) are conducting their survey at least partly by postal mail. All of them except one address the questionnaire directly to the person accountable for completing the form. 71% (10) enclose a covering letter, 57% (8) the results of the previous survey and 57% (8) a free return envelope (see figure 8).

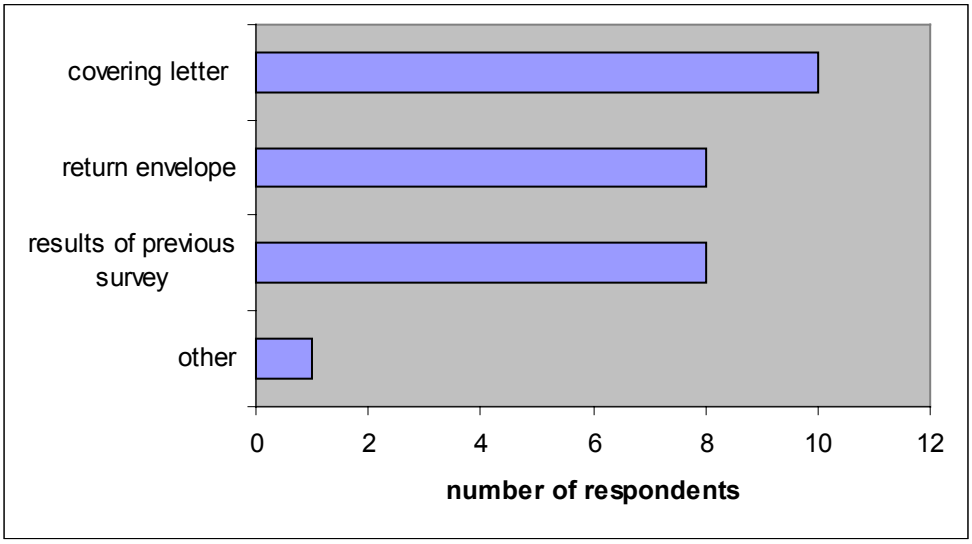
Four institutes conduct their survey with a mixed mode of postal/web survey. All send an announcement in advance, one by letter, the others by e-mail. Furthermore one institute provides the results of the previous survey online. At least one plan to conduct its surveys by web form.

Two institutes conduct a mixed mode of postal/e-mail survey, one country assists the mail part by an e-mail announcement, the other one confirms the receipt of the data.

One country conducts its survey with a mixed mode of interview/e-mail. The interviews are hold by phone while the results of the previous survey are handed out by e-mail.

The harmonised EU programme intends to ensure the comparability of the answers given to the BTS. In addition to these questions the institutes are free to ask extra questions. 87% (13) use this opportunity and add further questions to their surveys regularly. On average there are 12 supplementary questions per country and year with different periodicities.

figure 8: adds on in postal survey



source: question 3.3; several answers possible

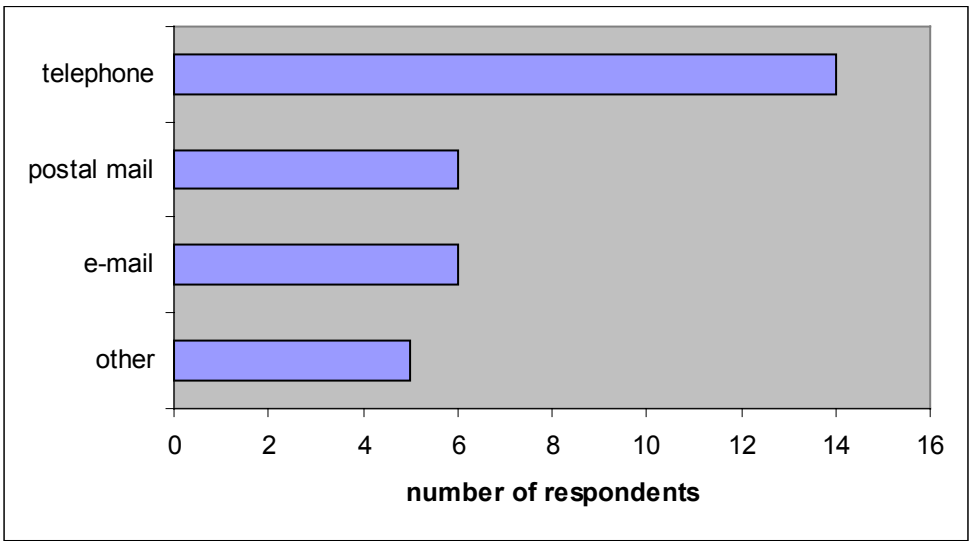
Every survey inevitably encounters the problem of imposing response burden in terms of time and effort for addressees. Besides others factors, the burden increases steadily with the number of questions. Nonetheless, countries in our survey asking additional questions did not show a deteriorated number of responses. Response rates in these countries exceeded the ones in the countries asking exclusively questions in line with the joint harmonised EU programme.

Hence participating may bring about a broad range of tangible and intangible benefits for the sampled unit stemming from the reputation of the surveying institute, pecuniary incentives or even civic virtue which surpass the respondent's cost of answering.

The mean sample size contains 1,909 units (1st quartile = 1,000; median = 1,700; third quartile = 3,000) and the institutes are aiming for an average response rate of 77%.

One of the most efficient measures to increase the response rate of the running wave are follow-up activities. This measure is applied by almost all responding institutes (14).

figure 9: follow-up mode

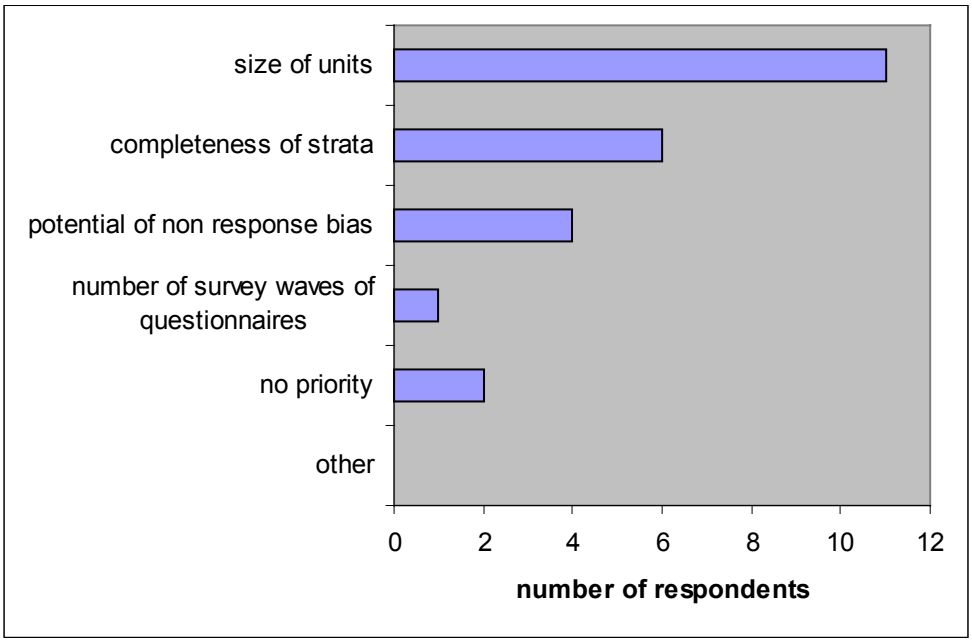


source: question 3.8; several answers possible

The different modes used for these activities are shown in figure 9. The category “other” has been specified in the following manner: four institutes are using fax, one of them visits the companies having given erratic replies. All institutes use at least telephone mode to get into contact with nonrespondents.

Altering the survey mode when contacting a nonresponding sample unit is likely to be constructive. According to our survey, the majority of the institutes which contact the units by a different mode in the follow-up phase exhibited considerable increases of response rates. Obviously a change in the mode increases the attention of the surveyed units. Moreover, the result could also be interpreted to suggest that it is desirable to provide a set of modes from which firms may chose periodically in order to ascertain that they are able to answer the questions in the most suitable fashion.

figure 10: priority of follow-ups



source: question 3.12; several answers possible

Budget and time constraints force the majority of the institutes to prioritise their follow-up activities. In figure 10 it is shown that “size of units” is the most reported priority, furthermore, it is even top priority for all who ticked more than just one option.

The average success rate of the follow-up activities differs quite considerably. It ranges from a minimum increase of 1.5 percentage points of the response rate to a maximum of 25. These activities were successful in 13% of the cases on average.

Despite all efforts to increase the response rate, some addressees do not return the completed questionnaires. But what about units who did not complete the questionnaire fully? Only 39% (5) of the answering institutes (13) have a minimum level below which a questionnaire is considered as unit nonresponse. But almost 57% (8 out of 14) are running a history of units’ nonresponses to supervise their non-answering behaviour, mostly for unit nonresponses.

The decision to continue to participate in the survey can be influenced to some extent by the respondent’s initial reasons for participating, such as interest in the survey topic, thinking the survey is important and so on, but also by expectation of a tangible reward. According to our survey 73% (11) of the institutes developed a reward scheme for participation. The reward given usually consists of the results of the former survey. Eight institutes are sending a general report and the remaining three are providing their respondent with general and individual reports. The participation to the survey are attractive for the companies as for 60% (9) of the reports it is the only way to access results.

All institutes are assuring confidentiality on their questionnaire. Confidentiality refers to the safeguarding of any information about one party that is known by another. As complete anonymity is virtually impossible, confidentiality is a real concern.

Rewarding participation should affect response rates positively. Measuring the immediate effect and size of the reward scheme, however, is tricky. The uncomfortable conclusion from our survey is that by simply comparing the mean response rate of countries with and without reward schemes we are unable to give evidence on the proposition. Response rates here are also affected by many other factors. Hence our results do not challenge the significance of reward schemes.

2.5 Missing data

In order to set up time series it is important to have a constant answering behaviour for all companies. Both bias and variance of survey estimates may be affected whenever a surveyed unit does not answer at all (unit nonresponse) or deliberately omits some questions (item nonresponse). In addition, a surveyed unit may decide not to answer either regularly or occasionally. In any case, the behaviour should be monitored as continually refusing to answer a question may contain some information. Furthermore inconsistent as well as illegible data is to be removed and has to be treated as missing data.

87% (13) check for item nonresponse, when questionnaires return. 73% (11) of the institutes verify the consistency of the values, whereas all check for invalid data in quantitative replies.

All organisations are confronted with the fact that some units do not answer in (at least) a single wave. Nevertheless, 60% (9) consider this problem in their evaluation. Three of them adjust weights, four imput the missing values with data of a similar unit and the remaining two institutes use a combination of these techniques. Six out of nine institutes can resort to their history of nonresponse.

53% (8) cope with regular nonresponse by replacing the unit with a unit of the frame list with similar characteristics.

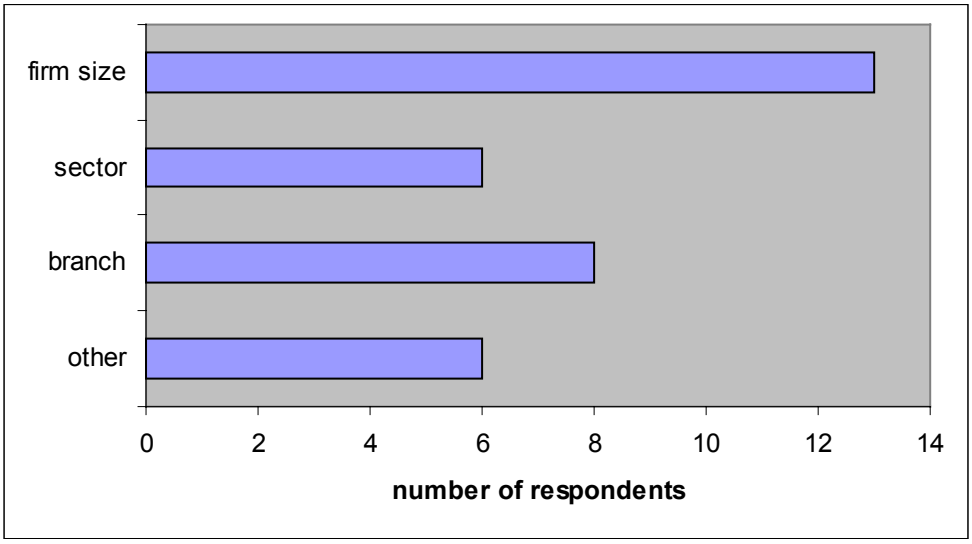
47% (7) of the institutes deal with item nonresponse. Three organisations adjust their weights, whereas three impute missing values. The follow-up activity of one institute is such intense that there is no item nonresponse at all.

2.6 Weighting

Sampling weights may improve the precision of estimates compared to a simple random sample under otherwise equal conditions. The Horwitz-Thompson estimator, for example, uses the inverses of the selection probabilities of the sampled units as sampling weights. If auxiliary information is obtainable which is related to the study variables at least to some degree, it can be used to further reduce sampling variance of the estimator. At the same time, auxiliary information can also be useful to tackle the problem of nonresponse and its potentially negative impact on bias and sampling variance of estimates. Sample-based (see question 3.14) and population-based weighting are common practice.

The use of auxiliary information for weighting is prevalent in the participating organisations. The institutes have reported utilising auxiliary information for weighting in some form, except one. Figure 11 shows the most frequently ticked options. Replies suggest that firm size comprises turnover data and/or employment size groups. A third of the organisations use firms' data on branch, sector and size for weighting.

figure 11: auxiliary information for weighting



source: question 5.1; several answers possible

In question 5.1 the institutes were asked to give information on their method of weighting when aggregating. At the first step, each reply is usually weighted in relation to the proportion of the firm's industry/employment size or turnover data. In subsequent steps, the aggregation process is mainly based on total value added or structural data of the economy. A more detailed examination of the aggregation is explored for a number of selected countries in section 4.

2.7 Quality assurance framework

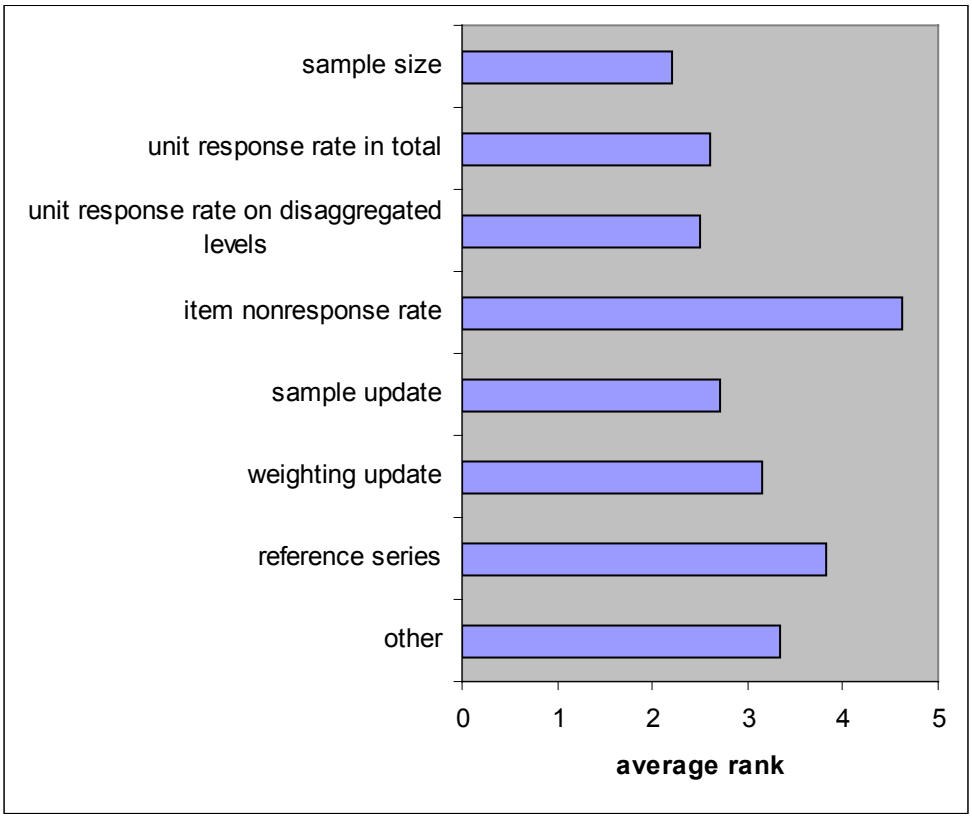
There is a growing demand for institutes to provide quality measures of their output. The source of the demand for the measurement and reporting of quality can be tracked down: On the one hand users want to judge the "fitness for purpose" of a statistical output, and on the other hand producers want to monitor the quality of their processes and products. Formulating comprehensive quality guidelines, however, is not trivial. The multi-faceted nature of the concept of quality calls for a variety of measures. The basis for an extensive measurement and reporting framework are the "Data Quality Attributes" which provide the concepts that a range of quality measures should cover.[3] Section 6 of our questionnaire covered the following areas.

- keeping frame/weighting information up-to-date
- data handling activities
- nonresponse
- benchmarking and sensitivity analysis

Ranked on a scale from one to eight, with one being very important, unit response rate in total turned out to be very important for the participating organisations on average (see figure 12). Similarly, sample size and sample updates were rated very highly. On the other side of the spectrum, item nonresponse rate and weight updates were not identified as key factors in the quality assurance consideration, although, participants' opinion on the importance of the former and, in addition, on reference series diverged most clearly.

[3] See footnote 1.

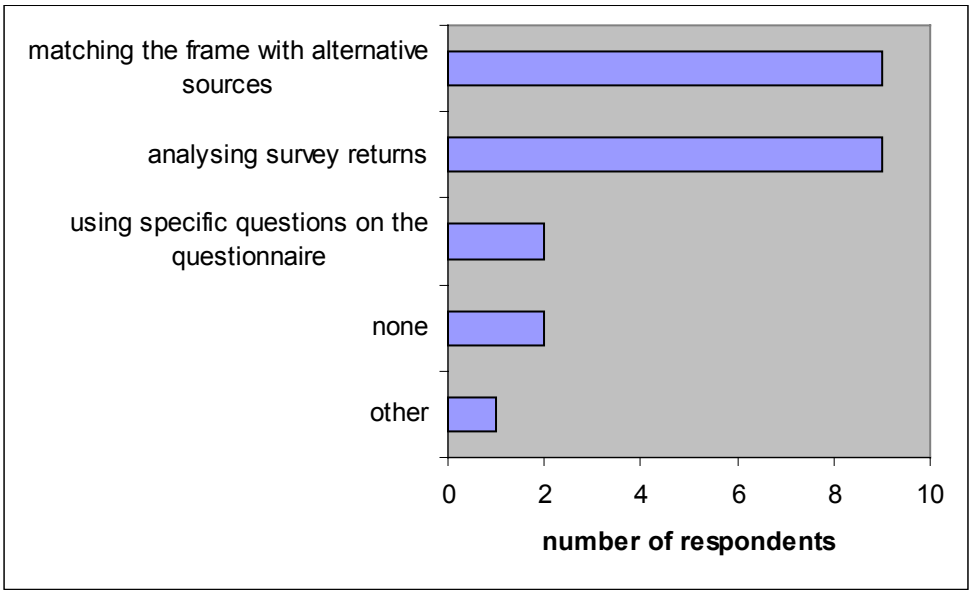
figure 12: key factors for quality assurance framework



source: question 6.8

Samples are mainly updated for changes in characteristics by both matching the frames with alternative sources and analysing survey returns. Only two of the organisations replied that they ignore any changes in characteristics of the sample. Two institutes use particular questions on the questionnaire to keep track (see figure 13).

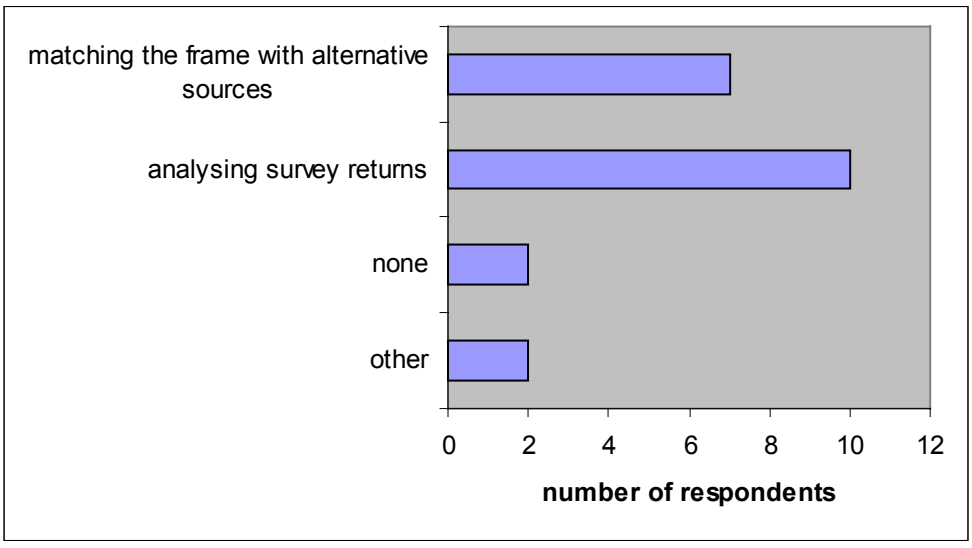
figure 13: update for changes in characteristics



source: question 6.1a; several answers possible

Similarly, matching the frames with alternative sources and analysing survey returns was reported as the preferred means of updating the frame list for exit from the frame population, albeit this is of limited use for handling entries (figure 14).

figure 14: update for entry/exit

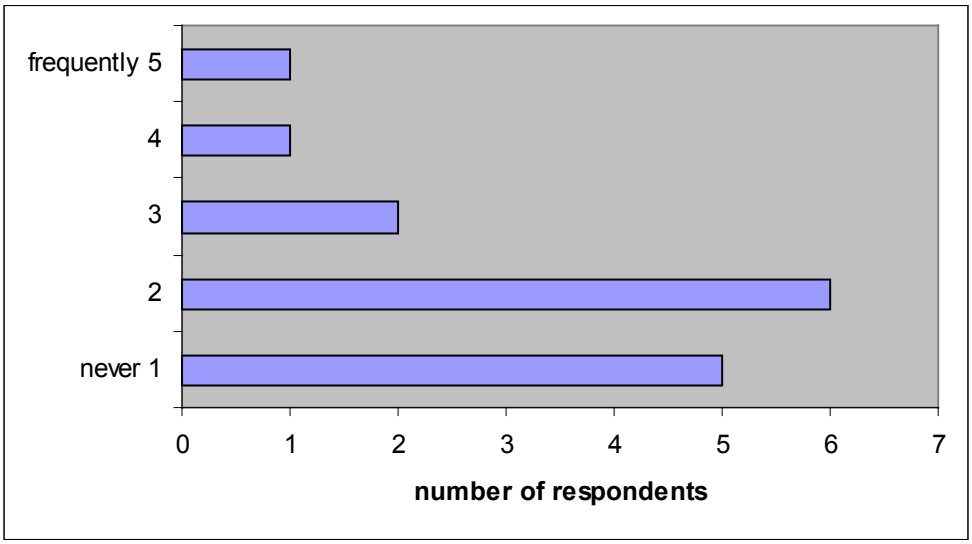


source: question 6.1b, several answers possible

Sensitivity analysis might be useful in consideration of the consequences of using alternative sampling methods or stratification options. It could involve speculation on benchmark scenarios and estimating the accuracy of data. As our survey shows, none of the participating countries actually addresses the issue of alternative sampling. But 60% (9) compare selected estimates to a benchmark series. Sensitivity

analysis to assess weighting factors is done by just two organisations. A single institute reported that it is calculating standard errors of computed estimates.

figure 15: nonresponse analysis



Source: question 6.3

Nonresponse, the decision of businesses not to take part in the survey, not only affects bias, but also sampling variance of estimates. As can be seen from figure 15, replies suggest that most of the participating institutes do not record and monitor item/unit nonresponse. Only six institutes collect this information.

Almost all of the organisations, however, make sure that replies on incoming questionnaires are checked in some form or the other. Question 6.4 considered how organisations manage the data processing phase. All of the respondents check completed questionnaires for invalid quantitative data (extreme values included), 87% reported that completed questionnaires are checked for item nonresponse and 71% reported checks for inconsistent data. To summarise the comments to that question, it can be said that organisations try to verify/correct replies with aid of official company sources or telephone calls.

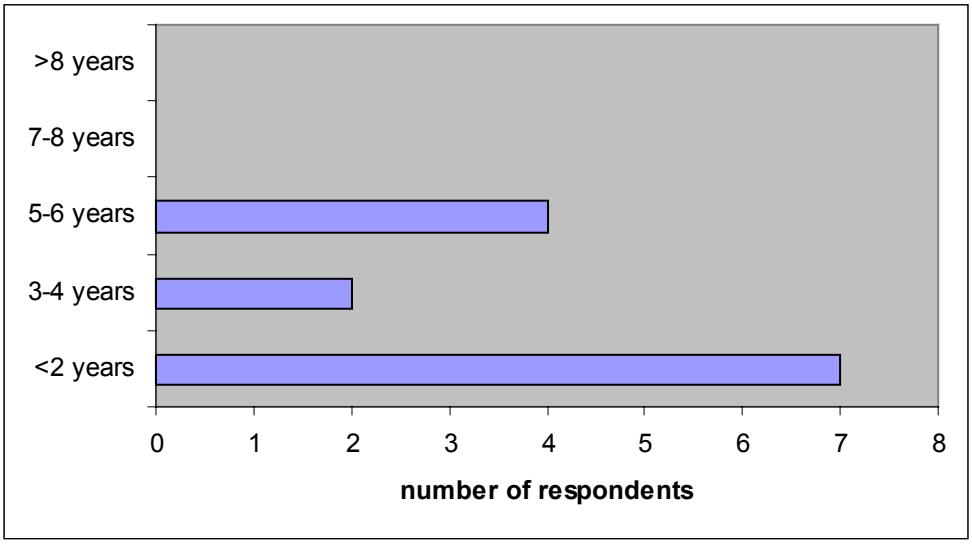
Critical values for response rates can be defined in respect to different aspects. Apart from the overall response rate, our survey focused on rates related to branches, strata and regions. In order to supervise the level of response, organisations put their effort mainly into reaching critical values for the overall response rate and the rates for branches and strata. Of course, this finding is partly reflecting the applied sample design and users' requirements for the survey, hence only two organisations consider response rates by regions as important. In general, response rates which are off the critical value are improved with follow-up activities by the institutes. After calculating

estimators, most countries check the plausibility of the results in regard to total, branches, strata and regions. Historical data, forecast data and alternative methods are used to assess the results. One country even has a professional department analysing and checking the data for integrity issues.

In question 6.5 of our survey, institutes were requested to give details on software. All organisations use various software packages to support the handling of the data. Most common are standard software packages which are partly enhanced by self-developed subroutines. Publicly available software to make specific survey calculations feasible (eg CLAN, GES, SUDAAN) was not mentioned by the institutes. This, however, could be linked to the fact that either generalised software is unlikely to be able to satisfy user's need or switching costs are to high. A large majority of institutes captures data from questionnaires with technical equipment.

Weighting is used to improve the quality of the sample according to the replies received (see question 6.7). Yet weighting factors are prone to become outdated whenever the fixed sample fails to map updates of the frame list instantaneously. In most of the participating countries, weighting factors are relatively recent. Seven out of thirteen organisations reported that weighting factors are no older than two years. There is no institute using factors older than six years.

figure 16: update of weighting factors



source: question 6.7

3 EU industry survey: effects of sample design

The purpose of this section is to examine some properties of the data from the EU industry survey. The analysis focuses on a set of time series of countries surveyed within the context of this project (except Ireland and Poland). After briefly describing the data for the empirical application, we report on some differences between the countries and try to give explanations in terms of sample design parameters.

3.1 Data set and descriptives

The data published by the Business Survey Unit of DG ECFIN are based on surveys carried out by public and private institutes. The industry survey is conducted on a monthly basis. In addition, it contains some questions asked on a quarterly basis (January, April, July, October). The harmonised EU programme encompasses seven monthly questions and nine quarterly questions. Total sample size in the EU currently amounts to about 35,000 units.

We do not analyse data for the full set of questions available but select them in order to cover the different types of question. First of all, we distinguish between monthly and quarterly questions. Because of the lower frequency and thus longer time period between observations, data on quarterly questions may contain more cyclical information (or a smaller share of random variation) than monthly questions. What is more, questions with different frequencies differ in their extent of how they react to economic shocks.

Within both monthly and quarterly questions, there are qualitative questions on past changes, on the assessment of the current situation and on expectations. Additionally, the quarterly questionnaire includes a multiple choice question (that will not be considered here) and quantitative questions.

We selected one question of each type for further analysis.

- Q1: Production trend observed in recent months (monthly)
- Q2: Assessment of order-book levels (monthly)
- Q5: Production expectations for the months ahead (monthly)
- Q9: Assessment of current production capacity (quarterly)
- Q11: New orders in recent months (quarterly)

- Q12: Export expectations for the months ahead (quarterly)
- Q13: Current level of capacity utilisation (quarterly)

For a majority of the countries data is available since January 1985. The data in this section range to May 2004 for questions one, two and five, and to April 2004 for the quarterly questions nine, eleven, twelve and thirteen. In addition, only seasonally adjusted data on aggregate net balances is considered.

3.2 Comparative overview by country

3.2.1 Published data

We start off by looking at the aggregated balance statistics for each question, as they are published by the European Commission. Table 2 below summarises some key statistics for both monthly and quarterly time series.

table 2: Some key statistics

Country	BE	DK	DE	EL	ES	FR	IT	LU	NL	AT	PT	FI	SE	UK
Sample Size	2000	500	3500	1070	2500	4000	4140	100	1700	1838	1000	680	1091	1000
Mean Q1	-0.5	10.9	-2.6	12.5	2.0	4.5	-7.8	-2.4	3.4	5.1	0.8	10.2	8.5	2.0
Std Dev Q1	7.0	11.6	9.8	9.1	14.3	15.6	11.6	20.1	4.6	7.7	8.1	15.0	18.8	17.1
Min Q1	-20.0	-21.0	-32.0	-19.0	-44.0	-38.0	-36.0	-75.0	-12.0	-28.0	-18.0	-28.0	-28.0	-46.0
Max Q1	16.0	41.0	22.0	36.0	33.0	36.0	17.0	79.0	14.0	29.0	19.0	47.0	48.0	35.0
Mean Q2	-20.3	-3.9	-21.1	-14.4	-15.7	-15.0	-7.0	-16.1	-7.9	-22.5	-16.0	-11.1	-24.3	-17.5
Std Dev Q2	12.4	16.4	16.0	8.1	17.7	19.9	12.4	24.0	7.3	14.6	13.4	23.7	23.2	19.1
Min Q2	-51.0	-41.0	-61.0	-32.0	-65.0	-64.0	-44.0	-64.0	-25.0	-57.0	-52.0	-70.0	-66.0	-62.0
Max Q2	11.0	32.0	12.0	6.0	11.0	27.0	16.0	34.0	5.0	11.0	8.0	39.0	25.0	26.0
Mean Q5	-3.4	9.8	1.5	24.8	5.5	5.6	9.7	-4.4	6.0	4.0	7.0	11.0	13.8	9.1
Std Dev Q5	10.8	7.6	8.8	7.1	7.4	10.8	10.5	16.4	4.0	8.3	7.8	12.6	16.8	14.5
Min Q5	-36.0	-23.0	-26.0	3.0	-19.0	-29.0	-20.0	-46.0	-4.0	-19.0	-11.0	-26.0	-19.0	-38.0
Max Q5	16.0	30.0	20.0	44.0	17.0	27.0	34.0	28.0	15.0	20.0	27.0	39.0	46.0	40.0
Mean Q9	27.4	15.5	17.3	13.0	6.9	17.6	26.8	17.5	4.7	13.7	11.8	20.1	6.5	28.3
Std Dev Q9	9.4	8.9	10.8	5.1	5.5	14.0	8.1	16.6	4.4	6.7	9.3	24.3	13.7	12.2
Min Q9	9.0	-5.0	-6.0	0.0	-2.0	-13.0	12.0	-12.0	-3.0	0.0	-4.0	-30.0	-19.0	1.0
Max Q9	54.0	33.0	42.0	25.0	20.0	46.0	43.0	57.0	16.0	22.0	40.0	81.0	30.0	55.0
Mean Q11	-3.6	7.3	0.9	12.6	5.4	-1.2	2.2	-6.0	6.9	0.7	-4.0	7.2	3.5	-1.3
Std Dev Q11	9.4	13.5	10.7	5.6	5.5	18.3	12.0	19.9	6.5	14.1	14.7	19.3	21.1	18.8
Min Q11	-24.0	-23.0	-31.0	-5.0	-8.0	-45.0	-22.0	-48.0	-12.0	-30.0	-37.0	-30.0	-37.0	-46.0
Max Q11	20.0	37.0	22.0	23.0	14.0	32.0	28.0	32.0	19.0	28.0	27.0	49.0	48.0	31.0
Mean Q12	-1.7	11.8	3.9	22.5	4.1	3.4	14.0	-7.3	7.6	5.4	9.1	13.6	19.1	4.3
Std Dev Q12	11.9	8.1	7.8	7.0	4.4	13.4	8.1	23.1	6.7	5.1	11.9	14.5	15.5	14.2
Min Q12	-26.0	-2.0	-16.0	6.0	-8.0	-34.0	-6.0	-58.0	-11.0	-8.0	-24.0	-21.0	-21.0	-34.0
Max Q12	35.0	24.0	17.0	41.0	12.0	28.0	31.0	41.0	22.0	15.0	30.0	44.0	49.0	30.0
Mean Q13	79.5	81.6	84.3	76.2	78.2	84.6	75.2	83.3	83.9	81.8	79.8	85.3	84.9	82.0
Std Dev Q13	2.4	1.8	2.6	1.7	2.3	2.4	3.8	3.1	1.4	1.8	2.3	2.5	1.6	2.6
Min Q13	74.5	76.4	78.3	71.5	71.1	78.5	73.7	77.8	80.3	79.1	72.6	80.7	82.2	77.4
Max Q13	84.8	86.1	89.5	79.3	81.8	89.1	80.8	89.2	86.2	85.0	84.3	89.9	88.5	86.4

source: own calculation, data taken from http://europa.eu.int/comm/economy_finance/indicators_en.htm. Country codes: BE Belgium, DK Denmark, DE Germany, EL Greece, ES Spain, FR France, IT Italy, LU Luxembourg, NL Netherlands, AT Austria, PL Poland, PT Portugal, FI Finland, SE Sweden, UK United Kingdom

Taking Q1, table 2 shows the fact that countries differ considerably in their value for the long-time average value (Mean) of the series. It is particularly noteworthy that the signs of the mean values vary although this property is not limited to this question.

As we mentioned in the preliminary remarks to this section, Q1 and Q2 differ in so far as the former is a series of changes, the latter one of levels. Table 1 readily shows that the standard deviation for the question asking for a qualitative assessment of changes is usually smaller than the one for the question which is related to some sort of a reference point defined in levels. Likewise, minimum and maximum values,

respectively, are usually larger for the series to Q2. This observation seems to be consistent with intuition as one would expect that survey participants experience movements within a short period less severe than the direct comparison of the current value of a variable to a somewhat more distant and fixed value.

When a sampled unit is asked to consider the current level of order books the survey deals with an individual judgement related to the level of the indicator. With this kind of questions, it is common that people assess the current level to a reference point (eg, normal level for the season as in Q2 of the programme) that could be some sort of an imaginary mean value. By this procedure the time series for Q2 is smoother than the data for series related to changes of some indicator. However, the amplitude of the time series is larger in Q2 than in both Q1 and Q5, respectively.

When looking at the table it sticks out that Luxembourg's data has by far the largest variance. At the same time, the country's sample size is the smallest of all which might partly explain the magnitude of the spread. However, to be fair, it has to be admitted that the variability of the data can be attributed predominantly to the pattern of the data until 1992. Afterwards, spreads have collapsed a great deal.

In Q5 survey participants are asked to give their expectations of future production trends. It is intuitively clear that through the process of expectation formation net balance data for Q5 should be smoother than, say, Q1, an assessment of encountered changes. Genuine uncertainty about future movements usually causes the fraction of stay-equal replies to increase indeed. Hence it is hardly surprising that comparing plots for Q1 and Q5 clearly shows that the graph for Q5 is much smoother than the one for Q1.

Comparing data for Q5 and Q1 it is immediately apparent that the curve is smoother than Q1's. As stated earlier this observation goes back to the smoothing effect of the expectation-forming process.

Explanation attempts for the quarterly data go along the same lines as for the monthly time series. Q11 deals with the individual judgement related to the level of a variable and thus the plotted graph is smoother than the on for Q9. The question on the degree of the capacity utilisation, undoubtedly, shows the smoothest graph. Of course, Q13 is special in that valid answers are restricted to positive values ([0,100]) by definition.

By doing some econometrics, it can be shown that data sets on balances to Q1, Q2, Q5, Q9, Q11, Q12 and Q13 are mostly (stochastically) stationary times series in many countries. This is absolutely in accordance with what we would theoretically

expect. However, there are time series for which a standard unit-root test clearly fails to reject the null hypothesis. These rejections may, for example, stem from structural breaks in the data caused by changes in a country's sample design. Apart from purely stochastic fluctuations, the (seasonally adjusted) time series of net balances also catch cyclical movements forecasters are typically interested in. But further analysing this problem would be beyond the scope of this study. Instead, we shall take a closer look at first-order differences of the survey data.

3.2.2 Differenced data

The behaviour of the published balance indicators only give limited information on the effects of the sample design. By running a differencing filter, business cycle effects should be removed. Hence variations are more likely to be attributable to sample design effects in a broad sense and randomness.

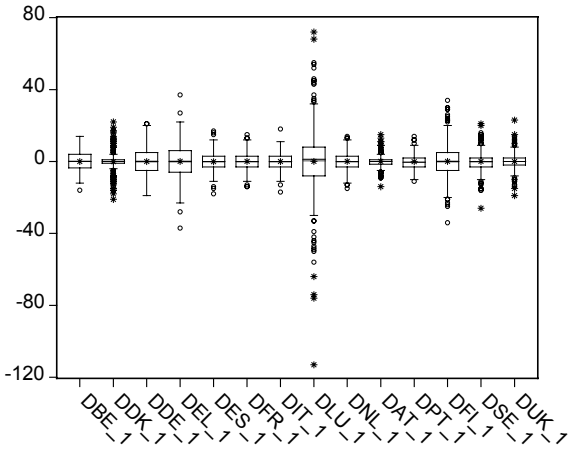
After differencing the data once, standard deviation for to Q1, Q2, Q5, Q9, Q11 and Q12 usually shrank considerably for the countries under consideration (see figure 17 for a graphical version of these data sets, in the form of box plots[4]). The box plots are based on the median and interquartile range which contains 50% of the values. Whiskers extend from the box to the highest and lowest values, excluding outliers. A line across the box indicates the median and the shaded area the confidence interval of the mean. Near and far outliers are distinguished with small circles and stars, respectively.

Figure 17 shows the summary plots based on the mean/median and spread of the data for the seven questions selected from the programme. What is immediately apparent is that time series from Luxembourg and – to a lesser degree – Finland stand out as having particularly large spreads. Denmark, for example, which has a comparable sample size to Finland does better in terms of spread in most cases. Therefore, only with sample size within the range of 100 units (Luxembourg) there is an indication of an increased volatility from period to period. In the other cases, there is no straightforward explanation for differences, indeed, but the fluctuation might, among other things, be caused by idiosyncrasies of the sample design of the countries' surveys.

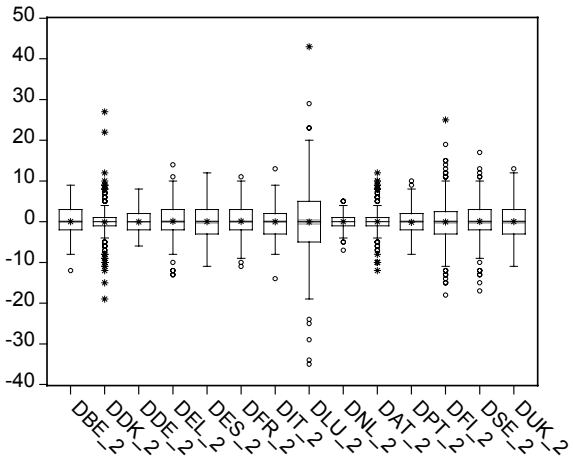
[4] Box plots are generated with EViews 5.0.

figure 17: box plots of selected indicators, first-order differences

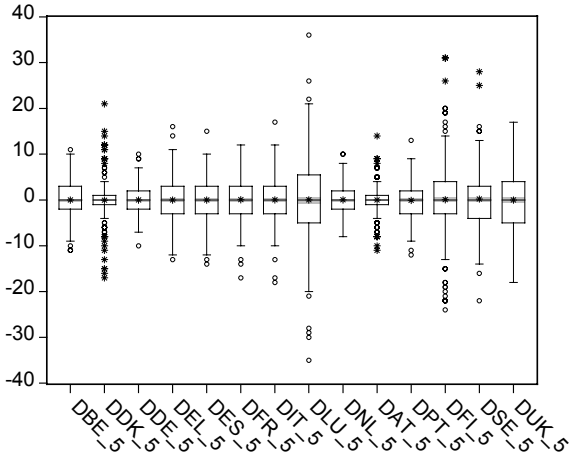
Q1: Production trend observed in recent months (monthly)



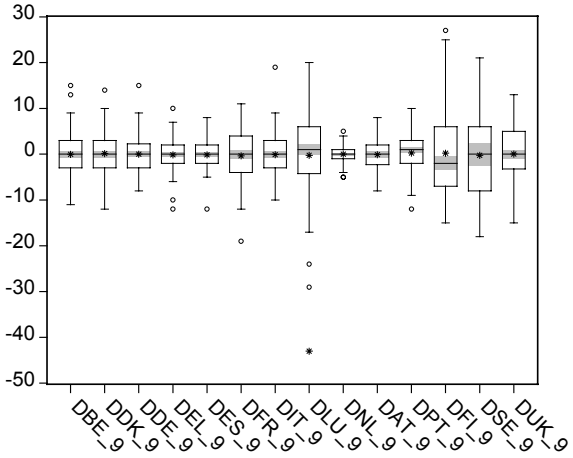
Q2: Assessment of order-book levels (monthly)



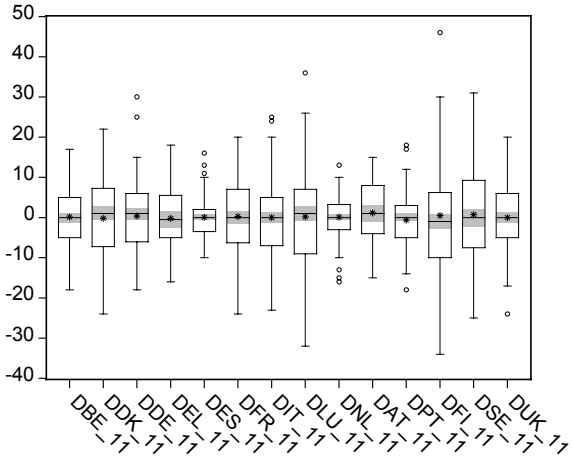
Q5: Production expectations for the months ahead (monthly)



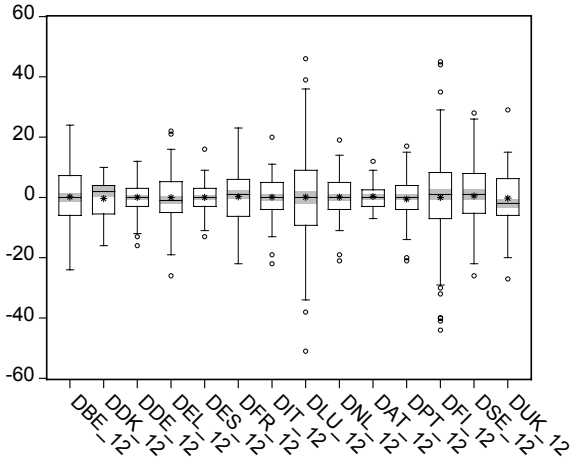
Q9: Assessment of current production capacity (quarterly)



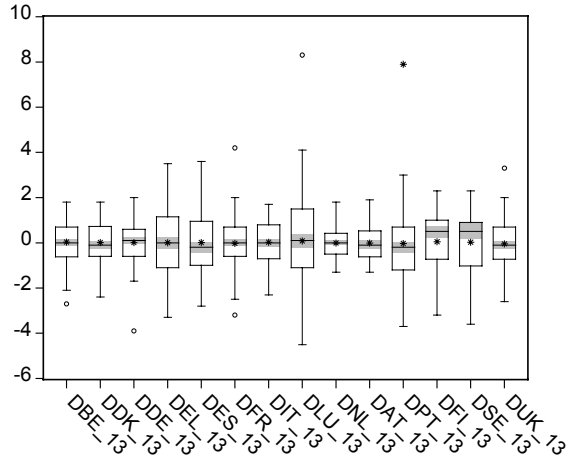
Q11: New orders in recent months (quarterly)



Q12: Export expectations for the months ahead (quarterly)



Q13: Current level of capacity utilisation (quarterly)



source: own calculation

Performing thorough econometric tests allowing to detect sample-design specific patterns is a complex and thus difficult task. The identification of these kind of effects in the net-balance time series would require an extensive analysis of the country-specific circumstances. For example, we related a country's sample size to the magnitude of its standard deviation for Q1, Q2, Q5, Q9, Q11 and Q12. A visual inspection of the scatter plots suggests that there could be a connection between the sample size and the standard deviation of a time series. Based on a strong theoretical ground, the corresponding regressions support the positive relationship in case of Q2 and Q5. However, the scope of the “detected” relationship is limited due to methodological shortcomings.

4 Analysis of selected surveys

4.1 Introduction

The selection of a few surveys in order to examine the sample design more thoroughly has to follow different aspects. First of all, the countries with the longest experience – the founders of the BTS in the present form – should clearly be included: France and Germany. In a second step, countries with some special weighting methods should be chosen. There is, on the one hand, Sweden which does no sampling and has different weights for different questions, and, on the other, the United Kingdom with its specific weighting scheme. In a third step, geographical and cultural aspects have to be considered. Central and northern countries are already selected. As a representative of the southern part of the EU Spain is considered. From the new member states joining the EU only few countries have a longer history in BTS and therefore a well established survey. For this reason, Poland was selected.

For each chosen institute, we look (see table 3) at the sources of the frame list, describe and assess the sampling methods and discuss the weighting schemes used. Moreover, the calculation of inferences are presented as far as it is done by the survey conducting institutions, and the treatment of missing data is considered. We identify the possible errors and break them down into non-sampling and sampling errors.

table 3: Overview, selected countries

	United Kingdom	Spain	France	Germany	Sweden	Poland
Frame List						
- Official data as sampling frame		X	X		X	X
- Missing sectors				X		
- Missing small firms	(X)		X	X	X	X
- Missing specific firms						X
Sampling Method						
- Stratification by size		X	X			X
- Simple random sampling	X	X	X			X
- Quota-sampling				X		
- No sampling					X	
- More than one questionnaire per unit	X		X	X		
Weighting						
a) Firm specific						
- Different weights for different activities			X	X	X	
- No Weights						X
- Employees	X	X		X		
- Value added						
- Question-specific			X		X	
b) Population oriented						
- Stratification by size		X	X	X		X
- Population weights on branch level			X	X		
- No additional weights					X	
- Value added	(X)	X		(X)		
- Employees			X			
- Turnover						X
Inference						
- Ad-hoc measures			X			
Missing Data						
- Weight adjustment	X	X			X	
- Imputation from previous waves			X	X		X
- Imputation from actual wave						X

() not completely

4.2 Frame list

4.2.1 General remarks

Compiling a frame list is easy – in principle. In practice, however, it can be a laborious task to cope with births of units, deaths, duplicates, organisational changes,

time delays, and mistakes, just to mention a few. Therefore, we first set out the definitions to define the background. In statistical terms a population is any entire collection of elements from which we may collect data. It is a basic concept as a sample is a set of items that have been drawn from the very population, say, all firms in the manufacturing sector. In general, however, the population from which a sample is taken, and the population of interest do not correspond entirely. The target population of a survey is not necessarily a population that can be surveyed. This gives rise to the notion of survey population. It is the actual population from which the survey sample is drawn. By studying the sample it is only possible to draw valid conclusions about the survey population, where some elements of the target population are deliberately excluded. For the sake of convenience and practicability a frame lists every element of the survey population together with basic identification variables making up the frame population. In reality the frame is affected by various imperfections. In this report the discrepancies between the target or survey population and the frame list are referred to as coverage errors.

The frame list is usually derived from a business register (BR), a comprehensive list of businesses that is used for statistical purposes. It is typically based on inputs from administrative sources and also provides the sampling frame for most business surveys. Investigation of its own and other surveys supplement these administrative sources. A BR typically consists of units (enterprises, legal units, administrative or tax unit, business location etc), a set of variables to each unit (NACE code, number of employees etc) and links to other units.

Updating procedures for units and variables are different and thus affect the quality of the BR information considerably. Because it takes time to correct the frame list the information on the frame population may be outdated and thus introduce frame errors. Bias caused by time delays in the information can be gauged after updates. Other errors may be never detected or only detected in the course of time.

Coverage errors, as mentioned above, are due to discrepancies between the target population and the frame population. They increase non-sampling error of the estimates and hence compromise accuracy. The most obvious case is, for example, the allocation of a unit classified in retail instead of manufacturing. The sources capable of deteriorating the accuracy of estimates can be categorised in two types: units belonging to the frame population but not to the target population, and *vice versa*. Under-coverage, the fact that a part of the target population does not occur on the frame list, introduces bias between the survey estimate and the true value. It could happen because of delays in entry registration. Conversely, over-coverage not only squanders scarce resources but also is likely to increase bias. It may emerge

from delays in exit registration or the problem of duplication of units. Unlike variance, bias is approximately invariant over repetitions of the survey. Coverage problems may also arise when working within the population, eg, with subpopulations.

In practical situations, there is at least some potential to confuse over-coverage and unit non-response. Units that appear wrongly on the frame list may have a completely different attitude towards the survey. If these units are less likely to complete questionnaires, the underlying coverage problem could be mistakenly interpreted as unit nonresponse.

To sum up, frame creation, maintenance and monitoring is an important part to improve and/or maintain the level of quality of business surveys.

4.2.2 The countries

Roughly half of the more closely examined surveys (in this section also referred to as (“the countries”) use a frame list consisting of data that is based on inputs from administrative sources, that is, France, Poland, Spain and Sweden. The CBI – conducting manufacturing surveys in the UK – uses its member register together with additional sources as, for example, company lists. The IFO in Germany also refers to companies enrolled for the survey. The predominant type among units are businesses defined as a legal unit, person or group of people. Sweden's National Institute of Economic Research, however, defines its units on a kind-of-activity level.

When compiling the frame list most countries deliberately exclude units from the target population not meeting certain listing criteria (eg micro companies, specific sectors, etc). Therefore the frame does not comprise all companies in the manufacturing industry. This fact has to be taken into account if making inferences. Thresholds on the number of employees and/or turnover size are the main discriminants. In the case of Germany, the whole energy and aerospace sector is not considered in its manufacturing survey. In the UK, the frame is not built up with data from the mandatory government survey register. The frame is based upon both CBI member and non-member companies instead and covers some approximately 30% of the employment within manufacturing. The frame lists of all the institutions are revised and updated at least every other year. The UK, Germany, France and Poland also regularly examine frame list entries for exit/entry, duplication and changes in characteristics of units.

From a statistical point of view, random sampling procedures produce defensible estimates of the frame population and sampling error. So, the definition of the population and its basic components is fundamental. The aggregation level of the

sampled unit is directly connected with the level relevant to the sampling frame. This means that whenever a random sample is drawn from a list composed on an enterprise level only the sampled enterprises are picked in a genuinely random draw. Problems may arise if a unit has two or more NACE activities and the survey is meant to give information on different NACE levels. Some institutes tackle this problem by sending a questionnaire to every site of activity related to manufacturing of the sampled enterprise.

4.3 Sampling methods

4.3.1 General remarks

It is usually not cost effective or practicable to collect and examine all the data that might be available from a population. Sampling provides a means of gaining information about the population without the need to examine the entire population. The quality, however, of the information from which the sample is to be drawn is crucial. In addition, the type of the sample needed has to take into account the type of analysis required. Basically, there are many different ways of how a sample can be drawn.

Total survey error may stem from two sources of errors: sampling and non-sampling error. Sampling error is an error resulting from using a specific sample which, compared to another sample selected by the same probability mechanism and assuming all other stages of the survey to be error-free, results in a different estimate of the study variable.

A common way to measure sampling error is sampling variance. In general, sampling error decreases with the sample size and depends on the size of the population and on its variability, it can be taken into account in the sample design and can be measured and controlled in probability sample surveys. One measure to evaluate a sampling method is to compare the actual variance with one obtained by a simple random sampling (design effect).

There are two different kinds of sampling methods used in order to draw a sample. Probability (equal and unequal probability) samples – the most common are simple random sampling, stratified sampling, cluster sampling and sampling with probability proportional to size – and non-probability samples – like quota sampling or voluntary sampling – can be distinguished. For non-probability samples it is impossible to calculate the probability of a unit being included in the sample, neither to estimate sampling variability nor to identify possible bias. The quality of the results can only be

monitored by comparing them to known values from the population. In contrast probability sampling is more complex and more expensive (in most cases) but without the previously mentioned disadvantages.

The basic form of probability sampling is simple random sampling. The frame list only has to contain identifiable population units. This sampling method does not need any auxiliary information, therefore the heterogeneity of the population can not be taken into account.

Stratified sampling is most common for large populations. The advantages are the following: reduction of variance, assessment of quality for each stratum, estimation for each stratum, coping with the problem of heterogeneity in the population. In fact, the more homogenous the strata are, the smaller the sample size of the strata. Therefore the stratification criteria have to be determined and the strata has to be designed (bounds and size). Usually the criteria – the most common are size and branch – are directly taken from the frame list. A post-stratification can also be applied.

4.3.2 The countries

Sweden uses all units of the frame list for their survey. A cut-off for firms with less than 100 employees allows to reduce the surveyed units to a manageable size. There is no answering burden for small firms and by the way the non-response problem is reduced but inference is not possible for the whole manufacturing industry. Nevertheless the weighting scheme is simplified as the inclusion weights are equal to one.

The UK draws its units by simple random sampling from their own register representing almost 30% of the industrial work force. Their sampling frame contains 2000 units. There is a mismatch between survey population and target population. The sampling method does not consider the variability of the population, but the sample contains mostly the largest companies. The replacement of a regular nonrespondent is hampered by the size of the frame list.

Poland, France and Spain use a mixed sampling method for drawing units. A stratified random sampling is used combined with an exhaustive survey of the largest companies. The stratification is based on the number of employees and the branch. Each of the three countries use over all of its branches the same bounds, the variance could be minimised by calculating individual bounds for each branch in the sense of Neyman allocation. Even if the variance decreases inversely with the square

of the numbers of strata[5], confining on three size classes is common for interpretational purposes and intuitive to communicate. The mentioned countries have cut-off points which are constant over all branches. The cut-off limits should be set in accordance with the distribution of each branch.

Quota sampling is similar to stratified sampling to the extent that the population is divided into different subpopulations. The main difference is that elements are drawn till the desired quota is reached. This sampling method is applied by Germany. The stratification criteria is branch and number of employee. The quota is almost 100% in strata for large companies corresponding to an exhaustive survey. The strata bounds are calculated for each branch, an evidence for an optimal allocation. The quota stratified sampling helps to recruit a sample of motivated participants, nevertheless it demands a permanent refreshing of the sampling scheme. The recruitment of new companies in the German sample is not based on a random sample but rather "volunteers" resulting from broad advertising campaigns. Therefore the quality measures of a probability sample can not be used, instead the estimates should be compared to the population.

4.4 Weighting procedure

4.4.1 General remarks

The manufacturing sector in most countries consists of such a high number of units that conducting a population census would be an overly expensive undertaking. To reduce the size of the population a portion of it is selected. There are principally two options: defining a cut-off limit and reducing with this measure the frame list to an acceptable size, or, determining the sample by any type of selection procedure. As we have seen, both methods are used by the survey conducting institutions.

Cut-off limits are used only for parts of small firms. Such a limit, eg firms with less than 20 employees are neglected, reduces heavily the frame list and in most cases the sample size. It reduces the burden of filling out regularly questionnaires for the small firms. Moreover smaller firms have often a lower initial response rate and the costs of recalls increase for the survey conducting institutions accordingly. The important disadvantage is, that the results give only the economic situation and development of the firms with 20 or more employees. But as results in different countries show, there is quite a distinct development between small and big firms. The inference from the big to the small firms is therefore not possible.

[5] Cochran, W. G., Sampling Techniques, Wiley, New York, 1977.

Despite the complexity a sophisticated weighting system can achieve, the main features of such a system may be sketched in the following way. The weighting of a study variable y consists in multiplying the observation y_i by a weight ω_i, for example, $\omega_i y_i$. The quantity $y_i^{(\omega)} := \omega_i y_i$ is thus the i^{th} weighted observation of the variable y.

The choice of the weighting factor and the weighting process take a great place in survey statistics. A good weighting of the data has to take into account of a lot of factors, which could significantly influence the results of the survey. In particular, the key elements are the design plan, the nonresponse and eventually a poststratification or some other adjustment to the whole population. For each of these elements, a weight is constructed, eg, $\omega_i^{(1)}$, $\omega_i^{(2)}$ and $\omega_i^{(3)}$ where typically $\omega_i^{(1)}$ is the inverse of the inclusion probability in the sample, $\omega_i^{(2)}$ is the inverse of the nonresponse rate and $\omega_i^{(3)}$ is the inverse of some relative measure of adjustment. It is customary to construct the final weight ω_i as the product of these different weights, ie, in this case as $\omega_i = \omega_i^{(1)}\omega_i^{(2)}\omega_i^{(3)}$. Of course, some fundamental aspects of the design plan such as a stratified sampling or a cluster sampling have to be considered in the weight ω_i and generally in the weighting process. Calibrating procedures can also be applied. Finally, we have to mention that for each study variable a specific weighting can be developed.

The auxiliary information caught in $\omega_i^{(3)}$ for the population-based weighting should have a close connection to the item to be questioned, otherwise the weights may introduce non-neglectable errors. For example for questions on employment, employment-related figures of the population should be used, for questions on production, production-related figures are demanded. Consequently, for questions on the assessment of order books (of stocks of finished products), the level of order books (of stocks of finished products)in the population should be available. But this information is normally not at hand. At some items therefore, an appropriate weighting does not exist and no weighting could probably represent the development better than the weights actually applied.

Independently from the population-based weighting, all institutions use an additional firm-specific weighting to adjust the answer to the importance of the unit's item. For example, a unit with a large production receives a high weight in a question on the production development. Correspondingly, questions referring to export activities could be weighted by a unit's share of total exports, ie, its export weight.

Unfortunately firm-specific weighting and population-based weighting go in the opposite direction[6]. In a stratified sample small firms have a lower probability to be selected than big firms and thus the strata of small firms get a higher sampling weight. But a very small firm within the strata of small firms get a lower firm-specific weight than a somewhat bigger firm in the same strata. If only two strata exist, this compensation effect is quite important. Using a high number of strata, the firm-specific weighting can almost be neglected.

In the subsequent section, the focus of the exploration will be to take a closer look at the countries' practices to aggregate replies. Whenever a specific probability sampling methods is chosen, $\omega_i^{(1)}$ is well defined and hence given by the method. Hence we shall not address the topic in this section. Neither shall we deal with the problem of weight adjustments owing to nonresponse. Instead we focus on the standard weighting procedure for aggregating survey replies.

4.4.2 The countries

Germany

The IFO uses as sampling frame based on non-administrative information. Quota sampling is used to determine the sample size. The units are firms, but if a firm is active in more than one sector, several questionnaires are delivered. Hence the sampled unit and the observation unit do not coincide necessarily. As a consequence, the firm-specific weight for the response of the kind-of-activity unit is reduced to the importance of the activity within the firm.

At the firm level, information on the number of employees of a firm or a sector of a firm is used. Weights do not go parallel with the number of employees but are trimmed down by weighting points. By doing this, the dominant effect of large employment units is scaled down. Whilst this method is likely to bias estimates, it can reduce mean squared error. This kind of weight adjustment reduces the effect of large firms and compensates for the lower probability of small firms to be selected.[7] On the other hand, there is no theoretically based way of determining scaling factors and thus the outcome on survey results should be monitored with care.

Replies are then weighted with respect to the relative size of the corresponding activity. Weights, however, do not alter with the type of the question. From a theoretical point of view, it could be beneficial to switch to a weighting scheme which adjusts weights in accordance with the question under consideration. Weighting is

[6] OECD, Business Tendency Surveys, A Handbook, 2003, page 37.
[7] Goldrian G. (Editor), ifo Beiträge zur Wirtschaftsforschung, München 2004, pages 26 et seqq.

based on the assumption that there is a close and direct relationship between the size measure and the study variable.

Although businesses with less than 20 employees and/or attributable to the energy or aerospace sector are entirely left out from the frame, statements from the survey are communicated as total German manufacturing trends. The IFO publishes more detailed data of survey results in respect of size class. Detailed data covers at least the categories small, midsized and large companies. The quantitative definition of stratum bounds on the branch level varies with industry sector.

In case of unit nonresponse, it is aimed to remove bias of estimates from failing units by a combination of weight adjustment and imputation.

France

The INSEE surveys all units exhaustively whenever the number of employees or turnover exceeds 499 or €150m, respectively. The frame population does not include companies having neither less than twenty employees nor less than €5m in turnover. This range of companies in manufacturing is left out.

Units in the survey are sent a questionnaire for every activity if possible. Hence the sampled unit and the observation unit do not coincide necessarily. When aggregating responses from the micro level, observations from the kind-of-activity unit are weighted by the relative weight in relation to an item-specific measure. This is, for example, data on a firm's export activity for export-related questions and so forth.

In order to decide on a firm's weight within size classes item-specific weights are calculated. Firm-specific weights are then used to compute survey results on branch level for all size classes characterised by employment. For the last step in the aggregation process, population-based weights come into play. During all steps, questions referring to the general assessment for future output or price are collected by using an arithmetic mean.

Survey results are always stratified by size, that is on the 3-digit and 2-digit, and French manufacturing level. The definition of stratum bounds are the same for each sector.

In case of unit nonresponse, imputation procedures are used in order to affect both bias and variance of estimates. The quality of the imputation method, however, dictates how effective this aim is achieved.

United Kingdom

For the UK's industrial trends survey of manufacturing, the CBI chooses random samples of companies from the frame list. Several questionnaires are delivered whenever a company operates in more than one NACE activities, and thus more than a single reply to every question is obtained from these companies. Register information of the company is used to weight responses of the kind-of-activity units. But no item-specific weighting is in place. In the following stages, the standard weighting mechanism is based on parameters of industry sector and employment. Data on gross value added broken down by size and sector is available for all manufacturing companies in the sample. Respondents are allocated into one of 200 groups defined by four employment size bands and fifty sectors.

Weights derived from gross value added broken down by size and sector are fixed over some period of time. Cell bounds in terms of employment size class and industry sector are identical regardless of sectors.

Because of the relatively small sample, large weights for some of the units could greatly influence the results. By the CBI's current weighting mechanism this problem is effectively alleviated. In case of unit nonresponse, it is aimed to remove bias of estimates from failing units by weight adjustments.

Sweden

The NIER carries out an exhaustive survey which is limited to companies exhibiting more than 100 employees. For companies on the list of Swedish businesses which consist of several kind-of-activity units several questionnaires are posted. All relevant information, eg value added, number of employees etc, are available on both company and kind-of-activity levels.

As all data relevant for weighting is available on the kind-of-activity level, the reported values are weighted on a kind-of-activity basis. Weights, however, are actual weights, as survey statistics are produced in the form of a census. In addition, there is item-specific weighting insofar as employment-related questions are weighted according to a unit’s number of employees. All other sort of questions are weighted uniformly by value-added data.

Larger firms with more than 100 employees are completely enumerated and surveyed which admits aggregating responses in a straightforward way. The fraction of businesses completely ignored by the survey might be considerable. As long as there is no clear evidence of a close connection between surveyed and non-eligible

units, making generalisations from the survey population's results is hardly methodologically sound.

The institute did not mention that it adjusts weights or imputes missing data in case of item and/or unit nonresponse.

Spain

The MCYT fully enumerates all units with more than 249 employee, the remainder is surveyed by a stratified sampling method. The unit on the frame list coincides with the observation/answering unit, that is, the selected firms receive a single questionnaire. During the first stage of aggregation, replies are weighted by firm specific employment data carrying over the same stratification criteria from the sampling procedure. The current procedure does not allow weights to change in relation to the type of question.

For the subsequent stages of aggregation survey results are weighted according to the relative importance of the sector or branch in terms of gross value added. In case of a single unit nonresponse, it is aimed to remove bias of estimates from failing units by weight adjustments. The institute does not replace regular unit nonresponse of a unit with another frame unit. In order to minimise bias and variance arising from item nonresponse, however, weights are adjusted accordingly.

Poland

The CSO addresses a single questionnaire to the selected unit of the frame list, that is, to each company. During the first step of aggregation responses are collected on a 2-digit level by three size bands. Size classes are fixed over industry sectors. Further aggregation processes use register turnover data to weight survey results. There are, to our knowledge, no item-specific weights. Weighting is not affected by adjustment procedures for missing data.

In case of a single unit nonresponse, it is aimed to remove bias of estimates from failing units by weight adjustments. The institute does not replace regular unit nonresponse of a unit with another frame unit. In addition, there are no adjustments for item nonresponse and thus attempts to minimise nonresponse error.

4.5 Inference

The question of inference in complex surveys raises several problems. First, assuming that one has found an estimator of a parameter, its statistical distribution

has to be specified. This task is generally done by assuming that the estimators are approximative normal distributed. This useful assumption, currently made in practice, permits to construct approximative tests and confidence intervals. These well known tools are necessary for a thorough and deep analysis of the estimates. But for this, one needs to compute the variance of the estimator and this is the major problem for the inference in survey statistics.

Indeed, in a statistical point of view, the publication of some point estimators is unsatisfactory if it is not accompanied with some indications of the quality of these estimators. In particular, it is always recommended to publish at the same time the variance of the point estimator.

At the present time, the question how to estimate in an adequate or optimal way the variance of the estimate of qualitative survey data has not found an answer. In computing the estimated variance of an estimator of some parameters (total, mean, median, percent, ratio, etc.), the survey statistician has to consider principally the following essential elements:

- the more or less complexity of the design plan (stratification, clustering, etc.);
- if a weighting scheme is applied;
- if some data are imputed or “multiple imputed”;
- if the estimator of the parameter of interest is a nonlinear function of the data.

All these elements lead to great empirical difficulties in estimating the variance of the estimator. In the majority of situations, there aren’t explicit formulas, analytically derived, which can be directly applied to compute the variances and some linearization or re-sampling methods have to be applied. Unfortunately, standard software will be no longer of great utility and ad hoc solutions have to be applied. Only the French institute INSEE declares to calculate standard errors, doing this with self developed routines.

Nevertheless, if we consider that the survey conducting agencies have to compute essentially only totals and means or proportions, that their design plans are basically a stratified sampling, it would not be a great task but very appreciated, that they compute and publish for these estimators an estimation of their variances. With some simplifying assumptions, one can relatively easily produce and program the relevant formulae. As instance, the well-known statistical software SAS, versions 8 and further, has incorporated as standard procedure the tools to do it. The handbook accompanying the software describes them very well.

If missing data have been imputed, computing the right variances or a good estimation of the variances is not so evident. In fact, sampling variability as well as the nonresponse model cannot be reflected. The problem could be simplified if a multiple imputation framework instead of a single imputation one is chosen. But, it seems that the commonly imputation method used by the survey conducting institutes is simply to reproduce the most recently reported value and, generally, none attempt to compute variances, and particularly variances taking into account imputed values, are made.

4.6 Missing data

4.6.1 General remarks

As sampling errors can be controlled by different sampling methods and optimal allocation, non-sampling errors are the bigger challenge to the survey conducting institutes in terms of total survey error. In general, an important component of non-sampling error is Nonresponse. There are different kinds of missing data the survey analyst is confronted with when conducting surveys: nonresponding behaviour leads to item nonresponse and unit nonresponse. Unit nonresponse occurs when there is no reply of a sampled unit, item nonresponse when a unit does not answer to a question. The recurring non-answering behaviour of a unit usually leads to its replacement in the sample from the frame list, a single wave nonresponse should be corrected.

Non-plausible, invalid as well as eligible values are removed after having been detected. They should also be replaced. Post-removed data should be treated in the same way as missing data. Missing data can deteriorate the quality of the estimates on the one hand as nonresponse may introduce bias and, on the other hand, the estimates are generally less precise as the number of respondents is decreasing.

In case of unit nonresponse it is common to cope with missing data by adjusting survey weights. In this case the weights of respondents are increased. For item missings (item nonresponse and post-removed items) imputing procedures are applied to missing data in order to fill in the gap. Both methods are able to decrease the bias and reduce variability of the estimates. There are different groups of imputation techniques that can be distinguished: mean substitution, single imputation (eg hot deck imputation) and multiple imputation. The first method consists in filling in the gap by using the mean of the values. The second method consists in selecting

values in a pool of similar donor units. The last method repeating the former procedure at least twice in order to maintain the variability.

4.6.2 The countries

Germany is coping with the problem of unit nonresponse by imputing missing values for large companies (recently reported values) and adapting weighting factors otherwise, item nonresponse is treated by imputation techniques. Sweden does not treat these problems at all, assuming that nonrespondents are behaving identically. The UK and Spain adjust survey weights for both types of nonresponse. Poland imputes by using the most recently reported value or using the values of a quite similar donor unit, they do not treat the problem of item nonresponse. France uses most recently reported values for unit nonresponse and scaling them if necessary (method of constant sample). Item nonresponse is corrected by imputation as well.

The design of the questionnaire has definitely an impact on item and unit nonresponse: the format and the length of the questionnaire, the survey mode, the survey time schedule, confidentiality aspects, legal authority and the reputation of the conducting institute and the sample design should be taken into consideration in order to conduct a survey in the most convenient way for the sampled unit.

All selected countries have additional questions to the harmonised EU programme. But there is no evidence for an impact on nonresponse.

4.7 Publications

4.7.1 General remarks

For advertising purposes and for rewarding activities it is indispensable to structure the monthly publications in the most comprehensible way. The sampling method, the aggregation method for qualitative and quantitative items as well as the estimates and their accuracy should be described to enable the comparability between countries. Besides general reports, individual reports can help to maintain response rate on a high level. In general, companies are interested in benchmarks to compare their situation with the branch’s mean. Smoothing factors and seasonal adjustment methods should be explained in a technical report which should be available.

4.7.2 The countries

Germany publishes the balances of the business situation of the German manufacturing industry as well as the business development and seasonally adjusted and smoothed time series. The specificity of Germany is taken into account by differentiating between estimates for the eastern and western part. Furthermore the German institution provides individualised reports which contain the results per branch. Some explanations for methodological devices can be found on the Internet, containing an example how the balances of the qualitative data are built. There is no note about the survey population, moreover the NACE code is only used for classification since April 2004.

The Swedish publication has an own appendix containing methodological notes. The stratified sampling as well as the aggregation methods are explained. An interpretation guide helps to understand the enclosed graphs. Some estimates are even published by branches in great detail with the response rate. Time series are seasonally adjusted, diagrams for subsectors are available on the Internet. The participants get an individualised report for their branch.

The French institution publishes information about their sampling and stratification method. Their weighting and aggregation methods are explained for the different items as well as for the strata. Non-weighted means are explicitly mentioned. All questions are qualitative, therefore the balances are given to the positive and the negative replies. The time series are seasonally adjusted. The results are published on a provisional basis, during the next wave they are revised to take late replies into consideration. Results for one sector are omitted to keep up confidentiality. The obligation to participate is mentioned.

The CBI publishes besides the trends review a technical guide to their survey. The institute encloses an exclusive survey brief to their participants with each questionnaire, containing broad-based current commentary, headlines and historical data for the activity in question. The technical structure of the weighting matrix is described. The transformation of qualitative data to the percentage balance statistic is given. The strict confidentiality guarantee is reinforced by mentioning the European Commission’s regulations on survey secrecy.

The Polish and the Spanish publications could not be analysed, as these institutions did not provide them in the required language.

5 Technical report

5.1 Sample design

Frame list

The distinction between population elements and sampling units is important by itself and also for the domains of estimation. Sampling units should provide initial coverage of the target population, ie, all sectors of business in manufacturing, small and large. Likewise, sampling units must allow valid coverage of the changing target/survey population as the initially selected sample ages. For a repeated survey like business tendency surveys, the sampling units need to be defined in a manner which allows the sample to be updated over time. That is, well-documented procedures to maintain and control the sample over time have to be implemented. These procedures should at least cover updating the sample for births, deaths, out-of-scope units and changes in characteristics. Moreover, auxiliary information, eg, employment data, used at a subsequent stage of the survey process should be kept as current as possible. Disregarding the information normally implies a bias, the inclusion, however, may imply higher variance. Basically, the frame list should include all businesses in the manufacturing industry, provided that a good frame for the whole population might be administered without excessive costs to providers. In any other case, the exclusion of part of the target population from the frame list should be kept at a minimum. Cut-off limits can help to avoid the burden of small companies and thus should be defined for each branch separately as the distributions differ among activity sectors. Cut-off limits and their effects on survey estimates should be monitored from time to time.

Most of the countries use official (business or government) registers as main source for their frame lists. This goes along with our recommendations. The remaining countries cannot use administrative information as there is no official list. A frame list should be considered that matches the characteristics of the target population as close as possible. Hence, the use of an official frame list would significantly improve the quality of the estimates.

It is supposed that some of the presently used cut-off limits are arguable. Countries neglecting a considerable part of the population or excluding whole branches should expand their frame list.

Sampling method

A sample drawn from the manufacturing industry should represent the surveyed population with its specific distribution. Two different types of sampling methods can be distinguished: probability and non-probability sampling. Probability sampling is to be favoured as it is the basis for statistical inferences. We recommend a stratified sample with simple random sampling within the strata. Stratification criteria are at least three size classes and industry branches. Size bounds should be determined by optimal stratification as well as the number of units to be drawn per strata. The method of optimal stratification even allows to draw units in certain classes exhaustively (for example all large companies). The share of the exhaustive strata can vary and nearly reach one for small countries.

Most of the institutes already act in accordance to our recommendations by using a stratified sample with simple random sampling within the strata. Institutes which dispose of official register data but use an alternative method, should adapt their sampling method accordingly.

Weighting procedure

Looking at the history of business tendency survey weighting methodology is inseparably linked to it whenever individual replies have to be put together to form a survey aggregate. Although there is always an issue about the explicit weighting scheme applied in a specific survey, there are some universal key elements a weight procedure should include.

- Data used for weighting should be as recent and accurate as possible.
- Firm-specific weighting should be based on the assumption that there is a close and direct relationship between the size measure and the study variable. For example, replies relating to units' expectations of future production should be weighted by value-added data instead of employment data if this information is available.
- The survey results should be compiled with respect to some stratification criteria.
- For aggregation, at least three size classes should be applied. The population-based weighting factors for the strata should be constructed from employment, value added or turnover.
- Stratification should not be limited to total manufacturing but on a branch level. The strata bounds have to vary between the branches.

- Population-based weights should be adjusted to suit the truncated population whenever the survey involves cut-off sampling.
- Sampling weights should be adjusted by constructing nonresponse weights to tackle the problem of unit nonresponse.

Most of the institutes stratifying their sample use in all branches fix bounds for their size classes. As the average size of companies varies between branches considerably, branch specific size bounds would improve the accuracy of the estimates.

Some countries apply question specific weights for the export related questions, but no institute uses specific weighting factor for the assessment of the stocks of finished products and the order books.

Inference

In order to have a clearer perception of the goodness of the estimates, it is necessary to publish them along with a measure of quality and precision. Thus the variances of the produced estimates ought to be computed as well. As Statistics Canada recommend in their quality guidelines a generalised estimation software should be used. Because procedures for estimation are mostly incorporated, the advantage of such a solution is to reduce time and cost for the survey analysts. Furthermore, computing the variance with a standard estimation procedure would be more transparent than using a tailor-made system.

Missing data

Even if there are different kinds of devices in order to reduce unit nonresponse like follow-up activities, it cannot be completely avoid. Doing nothing – due to lack of time – and using the estimates for the respondents is implicitly based on the assumption of identical characteristics between respondents and nonrespondents. From time to time this issue should be explored to discover potential bias from nonrespondent that has to be corrected.

There are several imputation methods, going from a simple donor unit to a multiple imputation. As BTS are usually conducted with a time restriction, the institutions should use the most suitable method for their computer software and their data set.

Only a few institutes analyse the nonresponse behaviour of units to discover a bias, the remaining institutes should explore the issue of nonresponse pattern from time to time.

5.2 Miscellaneous

Publications

The monthly publication is to be seen as a reward for the participating companies. Therefore it has to be tailored to the needs of the participants. Individual reports are preferred, at least for a specific branch activity. Furthermore the reader should be guided through the publication. Therefore the sampling method as well as the weighting scheme should be described. Aggregation methods as well as the weighting method should be documented for transparency and comparability reasons. The reader should understand the transformation of his item given to the published total. If confidentiality is jeopardised by publishing estimates of certain strata or sectors, they should be omitted.

Response rates, as well as applied imputation methods should be described. Methodological notes should inform the reader how his data had be treated until aggregation. This incremental proceeding is necessary for transparency.

Publishing time series, the institutions should explain their smoothing methods as well as their seasonal adjustments. A technical report containing formulae and technical devices should be available for further interested reader.

Quality control

The quality of the results depend heavily on the sample design. But a good design is worthless if its correct implementation is not controlled. Therefore, the following measures should be part of an efficient quality framework.

a) Quality control at each wave

- Treatment of missing or invalid data before the calculations
- Check for the response rate and to take additional measures if it is too low.
- Recalls within the month under consideration. Recalls should use a different mode to the mode of delivering the survey.

- Compare selected, aggregated BTS-data to benchmark series

b) Quality control in a regular time span

- Actualisation of the frame list within two years if data are available
- Control of the number of strata and its bounds
- Regular update of the weights according to the information on the population
- Calculation of the variance of the estimators
- Survey at the participating firms on the desired mode, the concept of the report, the design of the questionnaire

The majority of the institutes monitor the quality of their estimates relative to benchmarks. As BTS data, among other aims, are supposed to act as a proxy for quantitative indicators, benchmarking against manufacturing series builds an important part of the quality assurance framework and should be done regularly.

In order to control a constant quality estimates need to be analysed in terms of their sensitivity to crucial survey parameters as, eg, sample size, weighting, number of size classes, missing data etc. Until now this issue has hardly been treated by the institutes.

Documentation

A detailed written documentation should be available in a way that recalculations of earlier periods are possible. The documentation has to include:

- An exhaustive description of the sample design
- A history on questions and its exact wording
- The weights based on the population, not only the actual ones but on the history
- A history of the micro-data to all questions, including the firm-specific weights

Each institute should document how they conduct their BTS in manufacturing. Only a few institutes were able to send us this kind of documentation. In these cases they did not exceed the level of a general description also provided in the institutes' regular publication. The European Commission should initiate guidelines for the preparation of comprehensive manuals or handbooks, preferably in English.

Appendix

Questionnaire on sample design for business tendency surveys in manufacturing

Konjunkturforschungsstelle
Swiss Institute for
Business Cycle Research
ETH Zentrum WEH
CH- 8092 Zürich
kof@kof.gess.ethz.ch
www.kof.ethz.ch

KOF

Please complete this questionnaire, providing as much information as possible. If a question does not apply to you, please write NOT RELEVANT. Your information will be handled in confidence and used only in accordance with the European Union's data protection directive.

When completed please return this form by **31 March 2004 to:** KOF, Swiss Institute for Business Cycle Research, ECFIN/2003/A3-03, ETH Zentrum WEH, CH-8092 Zurich, Switzerland.
Should you require advice completing the form please e-mail the survey helpdesk: etter@kof.gess.ethz.ch.

WHEN COMPLETING THIS QUESTIONNAIRE PLEASE USE BLOCK CAPITALS.

Do not seperate the pages.

1. Sampling frame

1.1 What sources of information constitute the frame list of your business tendency survey (BTS) in manufacturing? (Please tick all that apply)

- ☐ business register
- ☐ government register
- ☐ other (please specify) .....
- ☐ own sources of information
- ☐ company lists

1.2 What is the size of the actual frame list? ☐☐☐☐☐☐

1.3 Does the frame list include all units of the manufacturing industry? yes ☐ no ☐

→ **If no**, what type of units are missing? (Please tick all that apply)

- ☐ specific branches (please specify) .....
- ☐ specific size categories (please specify cut-off points) .....
- ☐ specific regions/areas (please specify) .....
- ☐ other (please specify) .....

Konjunkturforschungsstelle
Swiss Institute for
Business Cycle Research
KOF

1.4 What characteristics does the frame list consider? (Please tick all that apply)

- ❑ identification
- ❑ address
- ❑ contact
- ❑ NACE code
- ❑ other (please specify) .....
- ❑ number of employees
- ❑ links between units
- ❑ links to other registers
- ❑ turnover

1.5 According to what periodicity is the frame list updated? (Please tick one only)

- ❑ 1 to 2 years
- ❑ 3 to 4 years
- ❑ 5 to 6 years
- ❑ > 6 years

1.6 Do you use any procedures to improve and/or maintain the level of quality of the frame list between the updates?

❑ yes ❑ no

→ **If yes,** do you have any procedures in operation with regard to

- ❑ the elimination of duplication
- ❑ the update for changes in characteristics
- ❑ the update for entry/exit of units

1.7 Do you deliver... (Please tick one only)

- ❑ only a single questionnaire to the whole unit
- ❑ a single questionnaire to the whole unit and if necessary several questionnaires for the activities in several branches

2. Sampling methods

2.1 How is the set of units selected from the frame list? (Please tick one only)

- ❑ simple random sampling
- ❑ stratified sampling with simple random sampling
- ❑ other, e.g. historical sample (please specify) .....
- ❑ sampling with probability proportional to size
- ❑ cluster sampling with simple random sampling

2.2a) If your organisation does use some form of stratified random sampling, please indicate the stratification criteria.

.....

.....

b) If you stratify by size on a branch level, please indicate the NACE depth (NACE rev 1).

.....

.....

2.3 Do you apply a rotation sampling scheme of the sample units? yes ❑ no ❑

→ **If yes**, does the scheme include

- ❑ all firms
- ❑ only a part of them (please specify)

.....

Konjunkturforschungsstelle
Swiss Institute for
Business Cycle Research
KOF

2.4 When deciding upon sample size, by what key factors is your decision guided?
(Please tick all that apply)

	not important				very important
level of precision of estimates	❑	❑	❑	❑	❑
amount of variability in the population	❑	❑	❑	❑	❑
budget constraints	❑	❑	❑	❑	❑
sampling fraction	❑	❑	❑	❑	❑
nonresponse	❑	❑	❑	❑	❑
other (please specify)	❑	❑	❑	❑	❑

.....

3. Response Rate

3.1 Is there a legal obligation for businesses to participate in the survey? yes ❑ no ❑

3.2 Are you running a sort of advertising for your survey? yes ❑ no ❑

→ **If yes**, please specify

.....

3.3 What kind of mode(s) are used in the manufacturing survey? (Please tick all that apply)

❑ interview mode → Do you...
- ❑ send an announcement
- ❑ fix up a date by phone
- ❑ other (please specify)

.....

❑ self-completion postal survey → Do you provide...
- ❑ a covering letter
- ❑ a free postal return envelope
- ❑ other (please specify)
- ❑ results of the previous survey

.....

❑ e-mail survey → Do you...
- ❑ send an e-mail announcement
- ❑ other (please specify)

.....

❑ web survey → Do you...
- ❑ send an e-mail announcement
- ❑ other (please specify)

.....

3.4 Who or what is the recipient of the questionnaire? (Please tick one only)

- ❑ a person if available (name, address, department etc.)
- ❑ department/company/job title (only unpersonalised)

Konjunkturforschungsstelle
Swiss Institute for
Business Cycle Research
KOF

3.5 Seven monthly and nine quarterly questions have been put forward in the joint harmonised EU programme of the manufacturing survey. Are respondents asked to answer additional questions?

yes ❑ no ❑

→ **If yes**

❑ regularly

❑ other frequencies

How many in 2003? ☐☐☐

3.6 What is the manufacturing survey's sample size? ☐☐☐☐☐

3.7 What unit response rate are you aiming for? ☐☐☐ %

3.8 Have you developed any follow-up activities? yes ❑ no ❑

→ **If yes**

❑ by telephone
❑ by postal mail
❑ by e-mail
❑ other (please specify) .....

3.9 What were the average response rates in:

	before follow-ups	after follow-ups
▪ 2001	☐☐☐ %	☐☐☐ %
▪ 2002	☐☐☐ %	☐☐☐ %
▪ 2003	☐☐☐ %	☐☐☐ %

3.10 What were the lowest/highest response rates after follow-ups in:

	highest	lowest
▪ 2001	☐☐☐ %	☐☐☐ %
▪ 2002	☐☐☐ %	☐☐☐ %
▪ 2003	☐☐☐ %	☐☐☐ %

3.11 Does the follow-up mode differ from the survey mode? yes ❑ no ❑

→ **If yes**, give details below .....

.....

3.12 How do you prioritise follow-ups? (Please tick all that apply)

❑ size of units
❑ completeness of strata
❑ potential of nonresponse bias
❑ number of survey-waves of missing questionnaires
❑ no priority
❑ other (please specify) .....

if several are ticked, specify which one has top priority .....

Konjunkturforschungsstelle
Swiss Institute for
Business Cycle Research
KOF

3.13 Is there a critical level for item nonresponse above which a unit is considered as unit nonresponse?

yes ❑ no ❑

→ **If yes**, please specify:

.....

3.14 Does your organisation have a history by sampling units of their nonresponses?

yes ❑ no ❑

→ **If yes**, does the scheme include

- ❑ Individual item nonresponse
- ❑ Individual unit nonresponse

3.15 Are you running an award scheme to maintain/increase participation level?

yes ❑ no ❑

→ **If yes**, by providing

- ❑ general report
- ❑ individualised report
- ❑ other (please specify) .....
- ❑ secured online connection to the BTS results

3.16 When providing results does your organisation make response rates publicly accessible?

yes ❑ no ❑

3.17 Is there a note on each questionnaire about the confidential treatment of the answers?

yes ❑ no ❑

4. Missing data

4.1a) Considering unit nonresponse, does your organisation deal with the problem of a single unit nonresponse by a respondent in a single wave?

yes ❑ no ❑

→ **If yes**

- ❑ weight adjustment
- ❑ imputation
- ❑ combination of both techniques
- ❑ other (please specify) .....

b) Does your organisation treat the problem of a regular unit nonresponse by replacement with another frame unit for future waves? (apart from regular updates)

yes ❑ no ❑

→ **If yes**, please specify

.....

Konjunkturforschungsstelle
Swiss Institute for
Business Cycle Research
KOF

4.2 Considering item nonresponse, does your organisation deal with the problem of item nonresponse?

yes ☐ no ☐

→ **If yes** (Please tick all that apply)

- ☐ sampling weight adjustment
- ☐ imputation
- ☐ combination of both techniques
- ☐ other (please specify) .....

5. Weighting

5.1 What kind of auxiliary information does your organisation use for weighting?

- ☐ firm size
- ☐ sector
- ☐ branch
- ☐ other (please specify) .....

5.2 During the process of aggregation, when does your organisation use weighting and how? Please provide details for each aggregation level.

.....

.....

.....

.....

6. Quality assurance framework

6.1 What kind of procedures are in operation to maintain the level of quality of the sample?

a) - update for changes in characteristic:

- ☐ matching the frame list with comparable alternative sources
- ☐ analysing survey returns
- ☐ using specific questions on the questionnaire
- ☐ none
- ☐ other (please specify)

.....

b) - update for entry/exit of units:

- ☐ matching the frame list with comparable alternative sources
- ☐ analysing survey returns
- ☐ none
- ☐ other (please specify)

.....

6.2 Are you using an alternative sampling method (e.g. simple random sampling) to evaluate sampling errors?

yes ☐ no ☐

→ **If yes**, please specify

.....

Konjunkturforschungsstelle
Swiss Institute for
Business Cycle Research
KOF

6.3 Differences between respondents and nonrespondents may be systematic and thus introduce bias. How often do you explore this issue?

never frequently
1 2 3 4 5
❑ ❑ ❑ ❑ ❑

6.4 Please answer the questions below considering the steps in the data processing phase.

- available data after editing returned questionnaires:
  - checking for item nonresponse yes ❑ no ❑
  - checking for invalid data (especially quantitative data) yes ❑ no ❑
    if yes, how do you deal with invalid data (please give details below)

    .....
  - checking for inconsistent data yes ❑ no ❑
    if yes, how do you deal with inconsistent data (please give details below)

    .....

- before calculating estimators:
  - checking the response rate in total yes ❑ no ❑
    if yes, how do you deal with the fact that the aspired value is not reached (please give details below)

    .....
  - checking the response rate by branches yes ❑ no ❑
    if yes, how do you deal with the fact that the minimum value is not reached (please give details below)

    .....
  - checking the response rate by strata yes ❑ no ❑
    if yes, how do you deal with the fact that the minimum value is not reached (please give details below)

    .....
  - checking the response rate by regions yes ❑ no ❑
    if yes, how do you deal with the fact that the minimum value is not reached (please give details below)

    .....

- assessment of estimators:
  - checking for plausibility of the results in total yes ❑ no ❑
    if yes, how do you deal with the fact that the values are not plausible (please give details below)

    .....
  - checking for plausibility by branches yes ❑ no ❑
    if yes, how do you deal with the fact that the values are not plausible (please give details below):

    .....
  - checking for plausibility by strata yes ❑ no ❑
    if yes, how do you deal with the fact that the values are not plausible (please give details below)

    .....
  - checking for plausibility by regions yes ❑ no ❑
    if yes, how do you deal with the fact that the values are not plausible (please give details below)

    .....

Konjunkturforschungsstelle
Swiss Institute for
Business Cycle Research
KOF

6.5 What kind of software tools do you employ? (Please tick all that apply and give some details)

- available data after editing completed questionnaires

❑ standard software packages .....

❑ self-developed routines .....

❑ other .....

- before calculating estimators

❑ standard software packages .....

❑ self-developed routines .....

❑ other .....

- assessment of estimators

❑ standard software packages .....

❑ self-developed routines .....

❑ other .....

6.6a) Do you compare selected values to a benchmark series?

yes ❑ no ❑

b) Do you apply any kind of sensitivity analysis to examine weighting factors?

yes ❑ no ❑

6.7 Considering the time lag between the update of the frame list and the update of the sample, how recent are the weighting factors:

❑ < 2 years ❑ 5 to 6 years ❑ > 8 years
❑ 3 to 4 years ❑ 7 to 8 years

6.8 What are the key factors which determine the design of the quality assurance framework, in order of importance on a scale of 1 to 8?
(1 = very important)

__ sample size
__ unit response rate in total
__ unit response rate on desegregated levels
__ item nonresponse rate
__ sample updates
__ weighting updates
__ reference series
__ other (please specify).....

6.9 Considering the quality of the computed estimators, do you calculate standard error?

yes ❑ no ❑

→ **If yes**, please give details on method and software used

.....

.....

Konjunkturforschungsstelle
Swiss Institute for
Business Cycle Research
KOF

Miscellaneous

Name, job title and address of the person who filled in the form: .....

.....

Telephone number: .....

E-mail: .....

Are there any comments you would like to make on this survey? .....

.....

.....

.....

We would appreciate it if you could enclose the following items:

- a copy of your BTS questionnaire(s) used in the manufacturing survey
- a copy of a BTS manufacturing report or any other publication containing the results
- a copy of your handbook, manual or documentation for the BTS in manufacturing if available (English, French or German)

*** **Thank you very much for taking the time to complete this questionnaire.** ***

Please return this questionnaire by **31 March 2004 to:**

KOF
Swiss Institute for
Business Cycle Research
ECFIN/2003/A3-03
ETH Zentrum WEH
CH-8092 Zurich
Switzerland